VISIT…

LANZAROTE

Caliente.com



面向21世纪的情报语言学

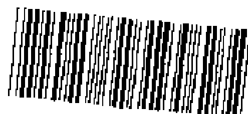
戴维民　赵建华　汪东波　主编

全国第三次
情报检索语言发展方向研讨会
论文集



北京图书馆出版社

全国第三次情报检索语言发展方向研


537780

面向21世纪的情报语言学

戴维民　赵建华　汪东波　主编

北京图书馆出版社

图书在版编目(CIP)数据

面向21世纪的情报语言学/戴维民等主编.—北京:北京图书馆出版社,2000.3
ISBN 7-5013-1693-7

Ⅰ.面... Ⅱ.戴... Ⅲ.情报语言-语言学
Ⅳ.H087

中国版本图书馆CIP数据核字(2000)第02780号

DM06/3602.

书名 面向21世纪的情报语言学
著者 戴维民 赵建华 汪东波 主编

出版发行 北京图书馆出版社(原书目文献出版社)
(100034 北京西城区文津街7号)
经销 新华书店
印刷 北京双桥威宁侯印刷厂

开本 787×1092毫米 1/16
印张 15.5
字数 360(千字)
版次 2000年3月第1版 2000年3月第1次印刷
印数 1-2000

书号 ISBN 7-5013-1693-7/G·455
定价 24.00元

前　　言

全国第三次情报检索语言发展方向研讨会经过一年时间的筹备,在全国情报语言学界的共同努力和支持下,于 1999 年 6 月 7～9 日在上海圆满召开。会议由中国图书馆学会、全国信息与文献标准化技术委员会、中国科技情报学会、全军军用主题词表编管会、中国索引学会、《中国图书馆分类法》编委会和空军政治学院共同举办,空军政治学院信息管理系承办。

在会议开幕式上,空军政治学院院长刘定昌到会祝贺,中国图书馆学会副理事长、国家图书馆副馆长、《中图法》编委会主任委员孙蓓欣,全国信息与文献标准化技术委员会主任委员、中国科技情报学会副理事长、中国科技信息研究所副所长梁战平,全军军用主题词表编管会委员、军事科学院百科研究部副部长马天保,中国索引学会常务副理事长葛永庆,空军政治学院副院长沈国权分别讲话和致词。

这次会议是继 1988 年在北京召开的"全国汉语叙词表发展方向研讨会"和 1996 年在北京召开的"全国分类法、叙词表发展学术研讨与成果展示会"之后,全国情报检索语言领域又一次规模较大的学术盛会,每次情报检索语言领域的学术盛会都标志着我国情报检索语言发展的一个新的起点、一个新时期的到来。1988 年的研讨会后,以《汉语主题词表》等词表的修订和大批综合性与专业性主题词表的出版为标志,我国情报检索用分类表、主题词表的研究与编制掀起了高潮,到 1996 年第二次研讨会,各单位编制的并实际应用的分类法、主题词表已达到百余部,除了继续维护与使用综合性分类法(如《中图法》)和综合性的主题词表(如《汉语主题词表》)外,又编制了部分多学科、多专业的分类法与主题词表,如《中国档案分类法》、《社会科学检索词表》、《军用主题词表》等实用性的标引检索工具。1996 年召开的第二次研讨会上,集中展示和演示了我国 90 年代前 6 年分类法与主题词表的成果,研讨了分类法、词表的编辑工作、提出了各种观点及技术问题,会议还提出了情报检索语言如何适应计算机检索环境和网络环境这一现代主题。

当人类即将迎来 21 世纪曙光的时候,共和国首先迎来 50 华诞。50 年来,作为新中国图书情报事业的组成部分的情报检索语言,作为图书馆学、情报学重要分支学科的情报语言学也同样取得了伟大成就,从建国初期开始对分类法理论和实践探讨,到 70 年代开始重视主题法的研究,表明我国作为情报检索保证的分类法和主题法已经得到高度的重视。一大批研究成果为一门新兴学科的创建作出了很厚实的积累。正是在这种积累的基础上,对各种分类法、主题法等进行抽象的、综合的分析与研究就构成了情报检索语言概念体系,作为一门新兴学科的情报语言学也就诞生了。情报语言学是研究情报检索语言及自然语言检索技术的一门学科。近 20 年的理论与实践成就推动着情报语言学不断走向发展,走向成熟。它的标志是:发表和出版了上万篇(种)的专业文献,形成了信报语言学基础著作和核心文献;涌现了一大批研究者,形成了一支优秀的核心研究者群,一些造诣很深的学科开拓者和学术带头人;建立了情报语言学学科理论体系和课程教学体系;编制了适应各类文献机构、各种文献类型的综合性或专业性的分类法、主题词表和分类法主题法一体化词表;制定和颁发了系列的检索语言编制标准、标引标准和作为标准的分类法和主题词表;在引进计算机技术,研究标引和检索技术等方面取得许多突破性进展,自然语言检索技术研究展现了良好的发展前景;对面向 21 世纪的新型情报检索语言和情报语言学开展了积极研究,取得了一些具有前瞻性的成果。

全国第三次情报检索语言发展方向研讨会总结了我国情报语言学理论与实践成就，深入研究情报语言学现实问题，探讨21世纪情报语言学发展方向，繁荣情报语言学学术研究。会议的主题是“面向21世纪的情报语言学”，会议重点研讨了中国情报检索语言的发展历史研究，情报检索语言的计算机化和电子化研究，自然语言情报检索技术研究，网络环境下情报检索语言研究，情报检索语言与标引技术标准化研究，中外情报检索语言比较研究，情报语言学系列课程建设和教学法研究，21世纪情报语言学发展趋势研究，情报语言学与索引工作现代化研究，情报语言学新理论、新方法和新技术研究等问题。

6月7日是会议开幕日，适逢张琪玉教授70华诞，为宣扬张琪玉教授执着的敬业精神，一丝不苟的严谨学风，空军政治学院举行了隆重的庆祝大会。张琪玉教授是我国当代著名的图书馆学家，情报语言学的开拓者和学术带头人。参加全国第三次情报检索语言发展方向研讨会的代表应邀参加了庆祝会，大会气氛隆重热烈，中国图书馆学会、上海图书馆和上海科技情报所、全军军用主题词表编管会、中国索引学会、北京大学信息管理系、华东师范大学信息学系等许多单位和个人发来了贺信，赠送了贺礼和花篮。在张琪玉教授70华诞前夕，北京图书馆出版社出版了《张琪玉情报语言学文集》，担任该书责任编辑的王燕来同志代表出版社为该书举行了首发式，并对这部汇集了张琪玉教授几十年情报语言学研究精华的著作给予了很高的评价。

这次会议共征集到论文40余篇，其中提交会议交流的有20篇，涉及情报语言学各个主要研究领域。大会安排了10人大会交流发言，分组进行了讨论，让与会代表都有畅所欲言的机会。陈树年总结指出，这次会议取得了许多共识，如：“要把受控语言与自然语言紧密的集合起来”，“情报检索语言要向易用化发展”，“有效的情报检索必定是受控的”等。

这次会议还集中演示了一些应用系统，有《军用主题词表》应用管理系统、《国防科技叙词表》电子版、分类表—叙词表转换系统、《解放军报》全文检索系统等，以及美国国会图书馆分类法机读版演示系统，这些系统展示了我国情报检索语言现代化技术进步和光明前景，同时也是对情报检索语言新理论、新方法的积极探索。陈树年在会议总结发言中说：“如果1996年第二次情报检索语言发展方向研讨会提出了受控语言与自然语言结合的方向，那么1999年的第三次情报检索语言发展方向研讨会则普遍确认了这个方向，并进一步提出了结合的方法和途径。这标志着我国情报检索语言的研究开始迈向一个新的阶段。”

在北京图书馆出版社的大力支持下，决定出版这次会议文集，在此我们表示衷心的感谢。

我们谨将此书献给在21世纪有志探索和研究情报语言学的同行们！

戴维民

一九九九年十月于上海

目　录

在“全国第三次情报检索语言发展方向研讨会”开幕式上的讲话……………………………孙蓓欣(1)
在“全国第三次情报检索语言发展方向研讨会”开幕式上的讲话……………………………梁战平(3)
在“全国第三次情报检索语言发展方向研讨会”开幕式上的讲话……………………………马天保(5)
在“全国第三次情报检索语言发展方向研讨会”开幕式上的讲话……………………………葛永庆(7)
在“全国第三次情报检索语言发展方向研讨会”开幕式上的讲话……………………………沈国权(8)
面向用户、面向新的知识载体、面向新的技术环境，发展我国21世纪的情报检索语言
——在全国第三次情报检索语言发展方向研讨会上的总结发言……………………陈树年(9)

中国情报检索语言五十年…………………………………………………………戴维民(13)
九十年代我国情报语言学研究进展………………………………………戴维民　傅　亮(50)
情报检索研究的认知观与情报检索语言的发展趋势……………………………胡建平(66)
浅析线性检索语言…………………………………………………………………陈　湘(70)

机读叙词表编制与维护技术的发展和未来…………………………………………曹　东(73)
《军用主题词表》应用管理系统……………………赵建华　赵建国　杨肖宁　吴　敏(82)
电子版叙词表及其编制技术………………………………………………………龚昌明(92)
计算机建立分类法和主题词或词串表转换系统的尝试…………侯汉清　李　波　戴晶萍(98)
分类号—主题词对照数据库的设计……………………………………张雪英　侯汉清(104)

军用主题词表系列化建设现状与发展……………………………………曹　杰　赵建国(116)
《军用主题词表》在文献自动标引检索服务系统中的运用…………………杜文芝　冯小平(123)
国防科技情报系统叙词语言的发展…………………………………………………龚昌明(127)
建立通信兵文献标引检索语言　实现通信兵文献管理标准化……………彭延生　田毅农(133)

论我国文献分类法产生和发展的基本规律…………………………………刘延章　苗　琳(137)
《中国图书馆分类法》发展的历史回顾与展望…………………………汪东波　卜书庆(142)
面向21世纪《中图法》及其标引技术改造的研究……………………………………陈树年(148)
文献分类法在网络资源组织中的应用………………………………………………马张华(165)
网络环境下分类法的变革与发展…………………………………………王忠红　张　涵(173)
图书分类法性能的定量测评初探——以《中图法》与《科图法》性能的定量测评为例
……………………………………………………………………………尚加宁　韩露盈(181)
主题词刍议……………………………………………………………………………文榕生(186)
增词标引与自由词标引刍议
——兼议《文献叙词标引规则》的有关规定……………………………………张云瑾(190)

我国自然语言检索研究的历史、现状及趋势 …………………………………………… 李新华(193)
积极为自然语言与情报检索语言的结合创造条件
——建议大量编制自然语言词表…………………………………………………… 张琪玉(199)
论网络环境下自然语言在情报检索中的地位与作用…………………………………… 黎盛荣(207)
《全国报刊索引》标引工作应用自然语言的设想………………………………………… 王越倩(211)
档案信息高速公路——情报检索语言在档案界发展的又一机遇……………………… 张正强(218)

开创情报语言学的新天地
——论张琪玉教授对情报语言学的新贡献………………………………… 曹树金　罗春荣(223)
张琪玉情报语言学学术思想简评…………………………………………………… 傅　亮(232)
《情报语言学基础》评介……………………………………………………………… 彭在曦(238)

在“全国第三次情报检索语言发展方向研讨会”开幕式上的讲话

中国图书馆学会副理事长
国家图书馆副馆长　　孙蓓欣
《中图法》编委会主任委员
（一九九九年六月七日）

各位领导，各位同行，同志们：

由中国图书馆学会、全国信息与文献标准化技术委员会、中国科技情报学会、全军军用主题词表编管会、中国索引学会、《中图法》编委会、空军政治学院等单位联合举办的“全国第三次情报检索语言发展方向研讨会”，今天在上海召开了，请允许我代表中国图书馆学会、国家图书馆、《中图法》编委会对这次大会表示热烈的祝贺。

这次大会是继1988年在北京召开的“全国汉语叙词表发展方向研讨会”和1996年在北京召开的“全国分类法、叙词表发展学术研讨与成果展示会”之后，全国情报检索语言领域又一次规模较大的学术盛会，每次情报检索语言领域的学术盛会都标志着我国情报检索语言发展的一个新的起点、一个新时期的到来。1988年的研讨会后，以《汉语主题词表》等词表的修订和大批综合性与专业性主题词表的出版为标志，我国情报检索用分类表、主题词表的研究与编制掀起了高潮，到1996年第二次研讨会，各单位编制并实际应用的分类法、主题词表已达到百余部，除了继续维护与使用综合性分类法（如《中图法》）和综合性的主题词表（如《汉语主题词表》）外，又编制了部分多学科、多专业的分类法与主题词表，（如《中国档案分类法》、《社会科学检索词表》、《军用主题词表》）。1996年召开的第二次研讨会上，集中展示和演示了我国九十年代前六年分类法与主题词表的成果，研讨了分类法、词表的编制工作、提出了各种观点及技术问题，会议还提出了情报检索语言如何适应计算机检索环境和网络环境这一现代主题。三年来，中国图书馆学会、中国科技情报学会、中国索引学会利用群众性学术团体的优势，积极组织所属会员开展该领域的学术研究，取得了与我国图书情报领域相媲美的学术成果；《中图法》编委会组织完成了《中图法》第四版的编制，完成了适应计算机检索需要的主题规范数据库及其维护管理系统，并正在研制《中图法》的机读版和电子版；《军用主题词表》编管会发挥军队优势，除规划本系统、本专业学术研究与词表编制外，还积极组织并参与全国性的学术活动，为我国情报检索语言的研究与应用作出了贡献；全国信息与文献标准化技术委员会在分类法、主题词表与文献标引领域标准的研制、修订方面作了许多工作，出台了系列国家标准；空军政治学院信息管理系作为我国情报检索语言领域学术研究最活跃的阵地之一，在开拓新理论、提出新观点方面成绩卓著。总之，我国图书情报界的广大同行已经充分认识到情报检索语言不论是在传统图书馆还是在网络环境下的现代图书馆，对于组织整序文献、开发利用信息资源都具有极其重要的作用。近几十年来理论与实践方面的成就推动着情报语言学不断发展、不断走向成熟，其标志是：发表和出版了上万篇（种）的专业文献，形成了情报语言学基础著作和核心文献；涌现了一支优秀的核心研究队伍，出现了一些造诣很深的学科开拓者和学术带头人；建立了情报语言学学科理论体系和课程教学体系；编制了适应各类文献机构、各种文献类型的综合性或专业性的分类法、主题

词表和分类法主题法一体化词表；制定了一系列的检索语言编制标准、标引标准和作为标准的分类法和主题词表；在引进计算机技术，研究标引和检索技术等方面取得了许多突破性进展，自然语言检索技术研究展现了良好的发展前景；对面向21世纪的新型情报检索语言和情报语言学开展了积极的研究，并已取得了一些具有前瞻性的成果。

今天，我们聚会在上海，召开"全国第三次情报检索语言发展方向研讨会"，目的是总结我国情报语言学理论与实践成就，深入研究情报语言学现实问题，探讨21世纪情报语言学发展方向，繁荣情报语言学学术研究。会议的主题是"面向21世纪的情报语言学"，会议将重点研讨中国情报检索语言的发展历史、情报检索语言的计算机化和电子化、自然语言情报检索技术、网络环境下情报检索语言、情报检索语言与标引技术标准化、中外情报检索语言比较、情报语言学系列课程建设和教学法、21世纪情报语言学发展趋势、情报语言学与索引工作现代化、情报语言学新理论、新方法和新技术等问题。我们希望，与会代表都能畅所欲言，充分发表学术观点、充分交流学术思想；集思广益，共同研讨我国情报检索语言领域的发展现状、发展中的经验教训和发展方向。中国图书馆学会、国家图书馆和《中图法》编委会将一如既往、责无旁贷地承担我们的义务，履行我们的职责，大力支持该领域的研究活动、开展该领域的研究工作。我们相信，通过大家的共同努力，一定能够进一步促进情报检索语言领域理论研究和实际应用的繁荣与发展，推动我国图书情报事业以崭新的姿态，迎接21世纪的到来。

最后，预祝会议圆满成功。谢谢大家！

在"全国第三次情报检索语言发展方向研讨会"开幕式上的讲话

全国信息与文献标准化技术委员会主任委员
中国科技情报学会副理事长　　　　　　　　梁战平
中国科技信息研究所副所长

（一九九九年六月七日）

同志们：

我很荣幸被邀请参加这次研讨会，今天，我代表全国信息与文献标准化技术委员会和中国科技情报学会讲几句话。

对于检索语言的研究，我国自50年代创建科技情报事业以来，就一直受到重视，40多年来，有一大批专家长期从事检索语言研究，在这方面获得了很大的发展，取得了卓越成绩。其中文献分类法就编制了130多部。比较有特色、影响面大，使用广泛的分类法有四部：中国人民大学图书馆编制的《中国人民大学图书馆图书分类法》、中国科学院图书馆编制的《中国科学院图书馆图书分类法》、北京图书馆和中国科技情报所合作编制的《中国图书馆图书分类法》和《中国图书资料分类法》。其中《中国图书馆图书分类法》已为85%以上的我国图书情报单位使用。

从60年代开始，我国又开始研究主题检索语言，共编制了主题词表120多部，其中比较有影响的大型主题词表有：中国科技情报所牵头组织全国编制的《汉语主题词表》，国防科工委信息中心编制的《国防科学技术主题词表》以及其它单位编制的航空、农业等专业主题词表。值得一提的是《军用主题词表》，从1987年开始编制以来，就作为一项实用系统工程来建设，投入了大量的人力、物力，完成了《军用主题词表》的系列化建设，为统一规范和繁荣发展全国的检索语言，推动文献数据库建设和提高全军文献信息传递、处理和资源共享的水平以及丰富和发展我国主题法理论做出了重要贡献。在此，我代表全国信息与文献标准化技术委员会和中国科技情报学会向《军用主题词表》编管会全体同仁表示感谢，感谢他们在检索语言及其标准化方面所作出的贡献。

80年代，我国又开始全面研究并尝试分类主题一体化的设想，其中代表性的成果有：《中图法》编委会组织编制的《中国分类主题词表》和中国社科院文献信息中心编制的《社会科学检索词表》。随着计算机在图书情报工作中的应用和近年来国际互联网的发展，我国又着手研究自然语言在检索中的应用以及在网络环境下检索语言的发展趋势，也取得一些重要成果。

在检索语言标准化方面，自1979年成立了全国文献工作标准化技术委员会（现更名为全国信息与文献标准化技术委员会），同时设立了检索语言分技术委员会，从此开始了我国检索语言标准化工作。

1980年10月，全国文献工作标准化技术委员会联合中国图书馆学会和中国科技情报学会在南京召开了全国分类法、主题法检索体系标准化会议，会议建议选用《中图法》作为国家标准分类法的试用本，取得了一定的成效。

1983年，我国第一个检索语言标准GB3860《文献主题标引规则》出台，现被修订为GB/T 3860－95《文献叙词标引规则》。以后，又陆续制定颁布了GB13190《汉语叙词表编制规则》、GB15417

《文献多语种叙词表编制规则》、GB15418《档案文献标引规则》等国家标准。可以说，到目前为止，建立了相对完整的检索语言标准化体系。

现在，随着信息技术和通信技术的发展，互联网伸向每个角落，信息量和信息种类大大增加，除了传统的文献信息外，还包括大量的图形、图像、语音、视频等电子信息。信息组织对象逐渐多样化，其范围也随之扩大，传统的信息组织方式远不能满足需要，其中占主导地位的分类法、主题法不可避免地受到挑战。未来的检索语言、特别是网络环境下的检索语言，将会具有机读化、国际通用化、兼容性、灵活性等特征，将会实质性地促进分类主题一体化的发展。目前，国际上较有影响的大型分类法和主题词表也都在这方面进行了积极的、卓有成效的尝试和探索，我国也要跟上这一发展趋势，使我国的检索语言及其标准化在网络环境下发挥更大的作用。我相信，这一次全国情报检索语言研讨会将一定取得圆满成功。谢谢！

在"全国第三次情报检索语言发展方向研讨会"开幕式上的讲话

《军用主题词表》编管会委员
军事科学院百科研究部副部长　马天保
（一九九九年六月七日）

各位领导、专家、来宾，同志们：

由中国图书馆学会、全国信息与文献标准化技术委员会、中国科技情报学会、全军军用主题词表编管会、中国索引学会、《中图法》编委会、空军政治学院等单位共同举办的全国第三次情报检索语言发展方向研讨会，今天隆重开幕了。我受全军军用主题词表编管会和军事科学院首长的委托，对这次会议的召开表示热烈地祝贺！向与会的各位领导、专家、来宾和全体同志致以亲切的问候和敬意！

情报检索语言研究和应用工作，是现代信息理论与技术的重要组成部分，也是文献信息现代化建设的基础工程。当前，以计算机和信息网络为主要特征的现代信息技术飞速发展，给情报检索语言的研究和应用提出了许多新的课题。作为检索语言工作者必须研究新情况，解决新问题，迎接新挑战，走改革创新之路。为总结和交流近年来我国情报检索语言研究和应用的新成果、新经验，深入探索未来21世纪现代信息技术条件下的情报检索语言发展方向，召开了这次学术研讨会。这次会议不仅对我国情报检索语言的繁荣发展具有重大意义，同时也将对我国文献信息现代化建设产生深远的影响。

全国情报检索语言发展方向研讨会首次举办是在1988年11月，第二次举办是在1996年7月，本次是第三次会议。从首次举办到本次会议，全军军用主题词表编管会每次都积极参与了会议的主办和筹备工作，而且受益匪浅，有力地推动了我军情报检索语言建设的发展。1988年以来，军用主题词表编管会组织领导全军圆满完成了我军第一部大型综合性词表——《军用主题词表》的编制和推广使用任务，并在此基础上完成了《军用主题词表》的配套工程和系列化建设，共推出各项科研成果30余项，荣获17项国家、军队科学技术进步奖。《军用主题词表》编管工作，对统一规范和繁荣发展我军的情报检索语言，推动各类文献数据库建设，提高全军网上信息资源共享水平，以及丰富和发展我国的主题法理论做出了重要贡献。我们深深感到，这些成绩的取得与我们军地之间的广泛学术交流和协作，特别是全国情报检索语言发展方向研讨会所产生的良好理论指导作用是分不开的。目前，军队的情报检索语言建设正处于由文本向电子化网络化的过渡和发展阶段，希望通过这次会议能够向与会的专家和同行们学到更多的新知识、新经验、新技术，为军队情报检索语言的发展注入新的生机和活力。

军民携手共同办会是全国情报检索语言发展方向研讨会的一个突出特点，这种办会形式不仅有利于军民之间学术交流和协作，而且有助于办会过程中各种实际困难的解决，提高会议的质量。希望今后通过各有关单位的共同努力，全国情报检索语言发展方向研讨会能够定期举办，越办越好，使其成为全国情报检索语言工作者的学术盛会和理论园地，不断把我国的情报检索语言科研事业推向前进。

这次会议在上海空军政治学院召开，空军政治学院的领导对这次会议非常重视，不但提供了非常好的会议环境，而且出席了今天的会议。空军政治学院信息管理系的同志们，会前做了大量艰苦细致的筹备工作，对会议的各项活动做出了周密的计划和安排。为此，我代表全军军用主题词表编管会和军事科学院领导对空政院领导和信息管理系全体同志表示衷心感谢。

最后，祝会议取得圆满成功。

在“全国第三次情报检索语言发展方向研讨会”开幕式上的讲话

中国索引学会常务副理事长　葛永庆

（一九九九年六月七日）

各位领导，各位代表：

“全国第三次情报检索语言发展方向研讨会”今天在上海揭开了序幕。来自全国各地的专家、学者和代表们，高兴地欢聚在浦江之滨，共同探讨和交流这方面的问题。借此机会，我代表总部设在上海的中国索引学会谨向会议的胜利召开致以衷心的祝贺，并向到会的全体代表表示热烈的欢迎。

随着图书情报事业的发展，作为情报语言学的主要研究对象——情报检索语言，近年来成果不少，在研究的深度上也有了很大的提高和进展。当然，我们也清醒地认识到，当前还面临信息泛滥的挑战，信息工作必须迅速提高质量，与此密切相关的情报检索问题因而显得更加突出，并变得相当复杂。为了要达到较高的检索效率，就必须作出很大的努力等等。因此，召开这样的研讨会就很及时而且十分必要，只有既加强理论研究的份量，又大胆创新，多出实践成果，才能适应形势发展的需要。

我们中国索引学会愿意与大家进一步加强联系，进行广泛而深入的交流，为做好世纪之交的信息工作共同努力。我们学会的副理事长张琪玉先生长期从事情报语言学研究，对情报语言学的理论建设作出了卓越的贡献。

最后，预祝会议取得圆满成功，并祝各位代表身体健康，在上海过得愉快。谢谢！

在"全国第三次情报检索语言发展方向研讨会"开幕式上的讲话

空军政治学院副院长　沈国权

（一九九九年六月七日）

各位代表，各位专家，同志们：

在即将跨入21世纪的前夕，共和国将迎来50周年庆典，中国图书馆学会也将迎来20周年纪念，在这个不平凡的岁月里，经学术界倡议，中国图书馆学会、全国信息与文献标准化技术委员会、中国科技情报学会、全军军用主题词表编管会、中国索引学会、《中图法》编委会和我们空军政治学院等七家共同主办的全国第三次情报检索语言发展方向研讨会今天在这里隆重召开，在此，我代表空军政治学院对会议的召开表示热烈的祝贺，向出席这次盛会的专家学者和与会全体代表表示热烈的欢迎。

世纪之交，信息在经济发展、社会进步中的作用与日俱增，如何快捷、有效地获取和利用各类信息越来越受到社会尤其是我们情报语言学界的广泛关注。随着计算机网络技术在情报检索领域的广泛应用，客观上对情报语言学的发展产生了极为深刻的影响，情报检索的计算机化给情报语言学的研究开辟了新的蹊径，拓展了众多崭新的领域。这次情报语言学界的同仁聚集申城，总结我国情报语言学理论和实践成就，深入研究情报语言学的现实问题，探讨21世纪情报语言学发展方向，必将进一步丰富和完善情报语言学的理论体系，增进情报语言学界的共识。这是本世纪末全国情报语言学专家的最后一次聚会，相信通过在座专家的共同努力，将推动情报语言学在21世纪走向更加成熟、更加繁荣。

这次这么多的专家学者汇聚上海，对我们来说是一次非常好的学习机会，特别希望各位专家对我院信息管理系的教学科研工作给予指导。

这次研讨会的准备工作，我们虽然作了最大的努力，但因条件有限，会务的组织保障一定存在诸多不周全之处，请各位代表谅解，欢迎提出宝贵意见。会议期间，我们还将穿插安排张琪玉教授70华诞庆祝会，会议代表联欢，看上海夜景，参观浦东，感受一下上海改革开放的气息，相信上海这座国际大都市日新月异的变化一定会给大家留下美好的印象。

最后祝各位代表身体健康，精神愉快，预祝研讨会圆满成功。谢谢大家。

面向用户、面向新的知识载体、面向新的技术环境，发展我国21世纪的情报检索语言

——在全国第三次情报检索语言发展方向研讨会上的总结发言

华东理工大学石油化工学院图书馆　陈树年

各位领导、各位代表：

"全国第三次情报检索语言发展方向研讨会"在热烈的气氛中，经过两天紧张、高效的研讨、交流、展示，即将圆满结束。会议达到了预期的目的。

这次会议是我国世纪之交的情报检索语言界的一次盛会，是在我国情报检索语言研究面临新的挑战和机遇、面临情报检索语言的理论研究与实践如何向纵深发展的重要关头召开的，因此对我国情报检索语言的发展方向具有重要的意义。

1　新形势及对情报检索语言的挑战

当前我们面临的新情况、新条件和新环境主要是：

(1)知识组织的范围不断扩大。已不仅是整本的文献、整篇的文献，而是逐步深入到微观的知识单元领域；不仅是传统的文献，而且已延伸到网上的各类信息资源。对更大范围的知识组织与检索、对网络信息资源的组织与检索，对情报检索语言组织知识的有效性提出了挑战。

(2)从文献情报整序到文献情报检索，所处的技术条件和环境发生了很大的变化。计算机技术、网络和通信技术、数字化和多媒体技术、超文本技术的广泛应用，要求情报检索语言自身的研究必须充分利用这些新的技术条件发展自己，而研究的成果必须适应于并服务于新的技术环境。

(3)情报检索用户发生了很大的变化。我们服务的已不仅仅是传统的图书馆读者和传统的情报用户，而是正在并且越来越多地向巨大的网络终端信息用户群扩展(有些尽管还是潜在的)，后者的情报需求多种多样，检索习惯与传统读者也有很大的差异。面对情报用户直接介入检索，要求新的情报检索语言应当是面向用户的，能满足用户更广泛地、快捷地、便利地检索情报的需要。

(4)随着网络时代的到来，我们比任何时候都更强烈地感受到信息爆炸对知识整序和信息检索的压力，用户对信息的时效性、有用性的要求越来越高。因此，情报检索语言必须能满足高速处理文献和有效过滤噪声的要求，自动编文摘、自动标引等技术必须有更大的进步，以做到迅速、准确、全面报道文献；另一方面，情报检索语言应有更好的动态性，能及时地更新、维护。这些都将推动人工智能技术在情报检索语言领域的应用。挑战和机遇是并存的，情报检索语言不但要适应新的检索需求和新的技术环境，努力改造、完善自己，还要积极地拓宽应用领域、主动出击攻占更多的"领地"。比如拥有浩如烟海信息资源的 Internet，本来是情报检索语言大显身手的舞台，但网上的众多"搜索引擎"，却是"非正规军"电脑爱好者的园地。我们应当认真研究网络环境下信息组织的特点和信息检索的特点，使情报检索语言在未来信息检索的主战场占有应用的地位。

2 对情报检索语言发展方向的共识

情报检索语言是我国图书馆学、情报学领域中非常具有活力的学科，是理论与实践结合得最紧密的学科之一。近20年来的硕果累累，形成一支具有很高水平的专家队伍和教师队伍，涌现出大量优秀的实践工作者，取得了一大批具有国际水平的研究成果。情报检索语言的生命力就在于不断适应新的条件和环境、不断满足新的知识组织和信息检索的需求。在第一组讨论中，张复华、鲍绵福提出，可以把"需求牵引，技术推动"的原则引入情报检索语言的研究，我也很同意这个观点，没有需求的牵引，情报检索语言研究就选不准方向，不借助现代高新技术的推动，情报检索语言的研究很难达到较高的水平，很难适应新的技术环境。我们这次会议的基本主题，正是紧紧扣住了这个关键。

从张琪玉教授的主旨发言到10位代表的大会发言，从提交的20多篇论文到各组的交流，大都集中在计算机和网络条件下情报检索语言的研究，这反映了代表共同关心的主题。代表们从不同的角度、不同的侧面、不同的检索语言出发，对21世纪情报检索语言的发展提出了很多相同的看法，如戴维民同志在《90年代我国情报语言学研究进展》一文中概括了网络信息环境下情报语言发展的动向，即情报语言必须具有相当高的兼容性；情报语言必须具有相当高的易用性；信息源的无政府状态迫使人们更改判断检索效果的概念和标准；数据源不仅在内容上而且在地理位置上都极度分散，情报语言对信息的组织功能应拓展到这一领域；情报检索需要研究多媒体信息特征表现和抽取的索引手段。

本次会议经过研讨、交流，在情报检索语言发展方向上取得了很多共识，这是本次会议取得的最重要的成果之一。其中第一个共识就是"要把受控语言与自然语言紧密的结合起来"。

张琪玉教授在他的《积极为自然语言与情报检索语言的结合创造条件》的论文中提出，将来的发展趋势是"情报检索语言的自然化和自然语言的情报检索语言化，是它们从互相结合到完全融合的过程"。

赵建华同志在《〈军用主题词表〉应用管理系统》论文中指出：对叙词表的研究，要置于计算机和网络环境下进行；叙词语言的创新，不是简单的机读化，而是充分与自然语言和计算机技术结合，克服自身存在的弊端，使之成为适应计算机和网络环境的新型叙词检索语言；增强自然语言与叙词的对应转换能力是叙词语言发展创新的关键环节，其方法是提高词表的入口率，对没有直接转换入口的自然语言要设法提供语义转换途径，增加叙词语言的自维护能力。

这些代表性的观点，反映了我国情报检索语言界在经受自然语言的冲击、经过近几年的探索后，对其发展方向取得的较为一致的认识。我认为今后新型的检索语言应当是"受控语言+自然语言+程序+通信网络+计算机"共同组成的一个系统，而不再会是停留在印刷形式上的词表或分类法。

另一个共识是"情报检索语言要向易用化的方向发展"。情报检索语言的易用化，既包括对标引人员的易用性，也包括对检索用户的易用性。易用化，重要的一方面还是情报检索语言自然语言化、智能化的过程。"傻瓜词表"概念的提出，就生动反映了情报检索用户对检索语言易用性的要求。易用化的程度越高，实际就是对情报检索语言的功能要求越高，只不过是把大量的词汇控制、转换等工作交给计算机后台处理，而交给用户的是直观、易学、易用、智能化的前台。

第三个共识是"有效的情报检索必定是受控的"。正如张琪玉教授所指出的"情报检索过程绝对不能没有控制，情报检索语言的控制原理将依然被保存，但将来的控制模式不再会是现在的控制

模式”。网络信息资源的迅速膨胀、全文数据库的出现、超文本技术的实现，使自然语言检索倍受青睐，传统的情报检索语言受到严重地挑战，于是有人认为“自然语言必将取代情报检索语言”，其实这是一种片面的认识。自然语言在语词检索、事实检索等方面的确有独特的功能，有利于快速开发数据库产品。但无控制的检索的代价是检索效率大大降低，试想同一个“计算机网络”的概念，一个是概括情报主题的标引词，一个是在经济文章中偶尔出现过一次的词语，尽管后者风马牛不相及，但同样也被检出，用户要花费大量的时间来筛选他所不需要的“信息垃圾”。宋明亮同志比较了我国使用无控制自然语言检索与使用受控语言和自然语言结合检索的几种全文报纸数据库的检索效率，前者检准率为 20% 左右，而后者为 80% 以上，相差之大，很能说明控制在情报检索中的意义。但传统的情报检索语言如不认真汲取自然语言的优点，以积极的态度向自然语言化方向发展，就适应不了新形势下情报检索的需要。

如果说 1996 年第二次情报检索语言发展方向研讨会提出了受控语言与自然语言结合的方向，那么 1999 年的第三次情报检索语言发展方向研讨会则普遍确认了这个方向，并进一步提出了结合的方法和途径。这标志着我国情报检索语言的研究开始迈向一个新的阶段。

3 情报检索语言研究成果的展示

本次会议不但进行了学术研讨和交流，还展示了一批研究的成果，如《军用主题词表》应用管理系统、《国防科技叙词表》电子版、分类表--叙词表转换系统、《解放军报》全文数据库系统、《LC》机读版的演示系统等。这些系统有的是代表我国最高水平的应用系统，有的是成熟的商品系统，有的则是试验系统，向我们展示了情报检索语言与自然语言相结合的新思路、新途径。而国外同行开发的系统则给我们以借鉴和启发。这些展示的成果，从一个侧面反映了我国当前情报检索语言研究的水平和动向。相对于叙词表的机读化电子化来说，我国分类表的机读化、电子化研究明显落后了；用于联机环境下的各类情报检索语言、具有与自然语言接口的实用检索系统、自动标引系统等的研制刚刚起步。我们期望在新的情报检索语言理论的支持下，积极开发研制满足新的知识组织与信息检索需求、适应新技术环境的情报检索语言系统、文献与信息标引系统和文献与信息检索系统，从面把我国图书情报工作的自动化水平提高到一个新的阶段。

4 建议与思考

会议期间代表们议论最多、感触最大的话题之一就是情报检索语言界缺乏有效的合作，包括合作的机制和相应的合作组织形式。呼吁情报检索语言界加强联系、加强合作，推动我国情报检索语言研究的快速、高水平发展。我认为我们的合作既应包括行业与地区间的合作，如图书、情报、档案系统的合作，教学、科研、企业、用户间的合作，军队和地方的合作，情报语言学与相关学科间的合作等；也应包括各种资源的合作，如智力资源、数据资源等。只有这样，我们才能做到优势互补，避免重复劳动和资源浪费，对重要的课题组织一批专家联合申请、协同攻关。但建立相应的合作机制和组织，还需要有关主管部门的积极支持。

有关情报检索语言的国家标准的修订和制订相对滞后，建议全国信息与文献技术标准化委员会加强在新技术环境下有关情报检索语言标准的研究与制订，以推动我国情报检索语言健康、规范地发展。

这次会议收到有关分类法研究的论文仅四篇，似乎分类法受到了冷落。现在我们谈“受控语

言"的自然语言化，主要还是指主题语言的自然语言化，但分类法在新的情报需求和新的技术环境下向何处发展似乎还比较困惑。相对于叙词语言而言，分类语言受自然语言的冲击更大，很多新兴的信息检索领域分类法很难插足，有些数据库甚至不进行分类标引。不知道是否是因为《中图法》一统天下的原因，以及分类法固有的"惰性"，在一定程度上制约了分类法的研究，不像主题法的研究在我国形成了百花齐放的活泼局面。分类法特有的系统检索功能是其他情报检索语言所不能代替的，应当说在用户由"提问一检索"向"浏览一检索"转变的今天，分类法也是可以大有作为的，关键在于研究与探索。分类法是典型的人工语言，是用代码表达主题的语言，不可能像叙词法那样直接地与自然语言融合，但它可以通过建立类号与自然语言的联系，实现相应的转换。在网络检索环境下，印刷型分类法对无数不可知的终端用户来说是无意义的，只有用层次简洁、结构明晰、适应网络信息组织的联机分类体系来控制信息的系统检索，才能为用户所接受。即使是对书目数据库来说，也必须极大地提高其动态主题描述能力和灵活的后组检索功能，才能适应新技术环境的要求。总之，分类法不是可有可无的、也不是无所作为的，对分类法的研究应引起大家的关注。

5 向张琪玉老师致敬

本次会议期间，时逢"张琪玉教授70华诞庆祝会及《张琪玉情报语言学文集》首发式"同时举行，全体代表参加了庆祝会。空军政治学院的院长、政委、各系领导，以及国家图书馆、中国科技信息研究所、上海图书馆、各学会等单位都到会祝贺，隆重而热烈的气氛，不但让张老师感动、学员感动，也让全体来宾深受感动。学院举办这样的庆祝会，不仅是对张老师的祝贺与褒奖，也是对广大学员的激励。此举充分体现了空军政治学院尊重知识、尊重人才的风气。

张琪玉老师在图书馆学情报学的实践、教学、科研战线辛勤耕耘40载，为我国图书情报事业的发展作出了重要的贡献。特别是他开拓了我国情报语言学研究的崭新领域，使我国在这一学科的研究走在世界的前列。张老师不仅自己的学术成果累累，还培养出一大批优秀的学生，其中不少人已成为我国这一研究领域的骨干，这些都是给我们留下的宝贵财富。张老师以其严谨的治学态度、不断创新的精神、紧密把理论与实践相结合的学风，成为我们学习的楷模。张老师在他的《钟情事业就是走向成功之路》一文中表示："科学需要一个人付出毕生的精力，这是最起码的条件，'衣带渐宽终不悔，为伊消得人憔悴'"。这种钟情事业的高尚情操，将会激励更多的人为祖国的图书情报事业献身的精神！

值此张琪玉教授70华诞之际，我们全体代表谨向您致以衷心的祝贺，祝您健康长寿、学术青春永驻！

这次会议的圆满召开，是与空军政治学院首长的高度重视和支持、信息管理系领导与教师的精心筹备和组织、蓝天宾馆工作人员的出色服务分不开的。对此代表们十分满意。再一次向空军政治学院的首长、信息管理系的师生、蓝天宾馆的工作人员表达我们的谢意！"蓝天如家兮情难去，白云作榻兮燕归来"，正表达了我们对这次会议和空军政治学院的美好记忆和留恋之情。

各位代表，这次会议就要结束了，相信这次会议将会把我国情报检索语言的研究与实践推向一个新的阶段。任重道远，让我们携起手来，满怀激情地迎接21世纪我国情报检索语言的春天！

中国情报检索语言五十年

空军政治学院信息管理系　戴维民

[摘要]本文对建国50年来我国情报检索语言的理论与实践作了全面的总结和回顾，以期浓缩我国情报检索语言精华。全文共分9篇，包括序篇、文献篇、人物篇、分类法篇、主题法篇、一体化词表篇、标准化篇、技术篇和未来篇。

[关键词]情报语言学　情报检索语言　中国　发展历史

1　序篇

当人类即将迎来21世纪曙光的时候，共和国首先迎来50华诞。50年来，共和国建设取得了辉煌的业绩，令世人瞩目。因此，总结50年的伟大成就，对21世纪的祖国建设和发展具有继往开来的重要意义。作为新中国图书情报事业组成部分的情报检索语言，作为图书馆学、情报学重要分支学科的情报语言学也同样取得了伟大成就，值得我们回顾和总结。

从建国初期开始对分类法理论和实践探讨，到70年代开始重视主题法的研究，表明我国作为情报检索保证的分类法和主题法已经得到高度的重视。但是我们也应清醒地看到，由于情报检索技术的落后，这些探讨和研究更多地还处于经验型阶段，尚未上升到理性阶段，用一个学科的眼光来审视这些研究，就可发现还没有构建起一个分支学科的基本框架。但是一大批学者的研究成果为一门新兴学科的创建作出了很厚实的积累。正是在这种积累的基础上，张琪玉开始对各种分类法、主题法等进行抽象的、综合的分析与研究，用情报检索语言这个极具创意的词汇统括了多种分类法、主题法及其他语言工具。1983年张琪玉《情报检索语言》一书由武汉大学出版社出版，使得情报检索语言一词更为广泛地应用和接受，在这一概念的统领下，学术研究空前的繁荣，加上我国情报检索语言编制实践也开始了新的跨越，这时，作为一门新兴学科的情报语言学也就应运而生了。学术界都在共同努力，丰富情报语言学的内涵。1987年《情报检索语言》一书易名《情报语言学基础》出版，对情报语言学作了较为准确的学科定位，它是研究情报检索语言及自然语言检索技术的一门学科。近20年的理论与实践成就推动着情报语言学不断走向发展，走向成熟。它的标志是：发表和出版了上万篇(种)的专业文献，形成了情报语言学基础著作和核心文献；涌现了一大批研究者，形成了一支优秀的核心研究者群，一些造诣很深的学科开拓者和学术带头人；建立了情报语言学学科理论体系和课程教学体系；编制了适应各类文献机构、各种文献类型的综合性或专业性的分类法、主题词表和分类法主题法一体化词表；制定和颁发了系列的检索语言编制标准、标引标准和作为标准的分类法和主题词表；在引进计算机技术，研究标引和检索技术等方面取得许多突破性进展，自然语言检索技术研究展现了良好的发展前景；对面向21世纪的新型情报检索语言和情报语言学开展了积极研究，取得了一些具有前瞻性的成果。

本文就是针对上述各方面的成就作一全面而简要的概述，限于篇幅不可能作深入的展开，但力

图从本文中可以看出我国情报检索语言理论与实践的发展脉络和主要成就，展望我国面向 21 世纪的情报检索语言和情报语言学的发展方向。

2 文献篇

情报语言学作为图书馆学、情报学一个比较成熟的分支学科，情报检索语言作为一个重要的研究对象，其中一个显著的标志就是发表和出版了一批数量可观的研究文献，其中不乏经典之作，形成了本学科的基础文献和核心文献。

2.1 文献年度分布

根据张琪玉编制的《情报语言学文献库》(电子版)按年度统计，其分布情况如表 2－1 所示：

表 2－1 我国情报语言学文献年度分布

年度	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	小计
文献量	54	0	1	5	0	11	8	9	13	62	55	218
年度	1960		1961	1962	1963	1964	1965	1966	1967	1968	1969	小计
文献量	39		6	1	7	30	78	49	10	0	0	220
年度	1970		1971	1972	1973	1974	1975	1976	1977	1978	1979	小计
文献量	0		0	3	42	7	19	22	10	36	19	158
年度	1980		1981	1982	1983	1984	1985	1986	1987	1988	1989	小计
文献量	101		265	292	333	462	507	603	590	525	682	4360
年度	1990		1990	1990	1990	1990	1990	1990	1990	1990	1990	小计
文献量	509		441	612	672	632	612	510	465	228	116	4397

资料来源：张琪玉：《情报语言学文献库》(电子版)，武汉大学出版社，1998

说明：1998、1999 年文献收录不全。

张琪玉编制的《情报语言学文献库》是目前我国收录情报语言学文献最齐全、数量最多的文献库，收录 1949—1999 年共计 9343 篇(种)，如果收全 1998、1999 年文献，50 年中总文献量可望达到万篇以上。与其他任何分支学科相比，其文献都是比较丰富的。

从统计表中可以看出，在 50 至 70 年代的 30 年中，情报语言学文献数量是很有限的，50 年代为 164 篇(种)，60 年代为 220 篇(种)，70 年代为 158 篇(种)，说明建国后的 30 年中，情报检索语言的研究是仅仅处于一个起步阶段，研究的问题面也很窄。特别是在建国之初和文革动乱的一些年度中出现了空白(1950、1953、1968、1969、1970、1971)，或文献量极少(1951 年 1 篇，1952 年 5 篇，1955 年 8 篇，1956 年 9 篇，1961 年 6 篇，1962 年 1 篇，1963 年 7 篇，1972 年 3 篇，1974 年 7 篇)的情况，说明建国之初百废待兴，文革中严重动乱，学术研究几乎处于停滞状态。

情报语言学作为一个分支学科，其发展和繁荣开始于 80 年代。80 年代前 5、6 年中，文献数量成百的递增，至 1985 年，年文献量已达到 507 篇(种)，整个 80 年代的文献量达到了 4360 篇(种)。

90 年代在 80 年代的基础上继续保持高度繁荣，在一个很高的平台上保持快速发展，在 1998、1999 年文献收录非常不全的统计下，总数也达到了 4397 篇，可见情报语言学在 90 年代已走向成熟，而且保持着强劲的发展势头。

仅从文献数量来看，我国情报语言学的学科发展历史不过 20 年。从 80 年代开始，情报检索语言从经验型和实用性的圈子中走出来，开始其学科理论的构造，搭建其学科框架，使情报语言学逐步走向完善和成熟。

2.2 文献主题分布

张琪玉和丘峰在对建国40年情报检索语言文献研究时，曾作过内容统计分析，虽然统计的数量有限，但其分布情况还大致能说明问题，如表2－2所示：

表2－2 建国四十年情报语言学文献主题分布

类别		篇数	百分比(%)
情报检索语言总论		159	5.5
分类法	分类法总论	413	14.4
	分类表体系、结构、编制和管理	197	6.9
	分类标引	529	18.4
	《中图法》	426	14.8
	中外分类法(《中图法》除外)	281	9.8
	小计	(1846)	(64.3)
主题法	主题法总论	142	4.9
	词表体系、结构、编制和管理	64	2.2
	主题标引	240	8.4
	《汉语主题词表》	45	1.6
	中外词表(《汉语主题词表》除外)	48	1.7
	小计	(593)	(18.8)
分类法和主题词法的比较和差异		61	2.1
著者号、书次法和目录排列法		191	6.7
情报检索语言人物和机构		75	2.6

资料来源：张琪玉，丘峰等著：我国情报检索语言40年文献简述，见《情报检索语言论文选》(书目文献出版社，1990年)

从上表可以看出，在建国后的40年中，我国情报检索语言研究主要是对分类检索语言的研究，占64.3%，同时可以看出主题法研究也日益兴起，已占了相当的比例(18.8%)。

根据张琪玉《情报语言学文献库》中的分类表进行浏览检索，文献量在50篇(种)以上的主题如表2－3所示：

表2－3 情报语言学文献主题分布

序号	主题名称	分类号	文献量
1	《中图法》评介与研究	Maa－Ne＝Dee	364
2	分类标引	Ei	250
3	情报语言学知识的普及	Ak	237
4	书次号	Eit	193
5	联机检索	Gsc	157
6	检索要求、检索效率	Be	155
7	情报语言学文献目录	Aa	136
8	藏书重新分类	Eix	136
9	数据库	Rn	134
10	检索策略	Et	123
11	著作评论	Age	118
12	主题目录编制法	Ekm	116
13	主题标引	Ek	104
14	光盘检索系统	Rp	99

序号	主题名称	分类号	文献量
15	分类法与主题法问题	Bgh	95
16	计算机情报检索系统	Rm	95
17	计算机情报检索	Gs	94
18	情报检索系统著作	Es	75
19	统计语言学标引法	Hk	72
20	与知识分类关系	Aeb	67
21	分类目录编制法	Eim	67
22	关键词法	He	62
23	文本检索	Hg	61
24	主题标引	Kg = Ek	61
25	某学科、某类型文献的分类标引	Eie	60
26	与语言学关系	Aed	59
27	主题索引	Fvb	57
28	《国际十进分类》评介和研究	Mhf-Ne	57
29	情报语言学术语	Af	53
30	《杜威十进分类法》评介和研究	Mhb-Ne	52
31	分类检索语言发展史	Bia-Ka	51
32	情报检索语言评价问题	Bqw	50
33	分类法主题法一体化	Bqi	50

资料来源:张琪玉:《情报语言学文献库》(电子版),武汉大学出版社,1998

上表反映了我国情报语言学研究的主要方面,实际上也展示出了情报语言学的研究重点。

2.3 论文情报源

根据统计,我国情报语言学论文的情报源如表 2－4 所示:

表 2－4 情报语言学论文情报源

刊　　名	发文量	刊　　名	发文量
图书馆学研究	344	高校图书馆工作	97
中国图书馆学报(含图书馆学通讯)	234	医学图书馆通讯	95
情报理论与实践	230	上海高校图书情报学刊	92
图书馆建设(含黑龙江图书馆)	219	图书馆工作	86
图书情报工作	218	图书馆工作与研究	80
科技情报工作	193	江西图书馆学刊	79
现代图书情报技术	183	北京图书馆通讯	75
图书情报知识	181	情报学刊	75
江苏图书馆学报	180	情报科学	73
图书馆杂志	178	河南图书馆学刊	66
大学图书馆学报	173	国外情报科学	57
图书与情报	159	青海图书馆	56
四川图书馆学报	143	图书馆杂志	46
图书分类论坛	137	湖北高校图书馆	45
图书馆理论与实践	137	资料工作通讯	44
山东图书馆季刊	126	四川图书馆	44
情报资料工作	122	图书馆论坛	43
情报科学技术	118	大学图书馆通讯	43
情报学报	111	高校文献信息学刊	42

刊　　名	发文量	刊　　名	发文量
云南图书馆	108	国外图书情报工作	41
情报业务研究	107	北京图书馆馆刊	38
图书馆学刊	103	江苏图书馆工作	37
图书情报论坛	102	宁夏图书馆通讯	29
文献工作研究	98	计算机与图书馆	29
图书馆研究与工作	97	图书馆	26
图书馆论坛	24		

资料来源:张琪玉:《情报语言学文献库》(电子版),武汉大学出版社,1998

上表展示了我国51种专业期刊登载情报语言学文献递减情况,可以确定登载100篇论文以上的刊物为情报语言学的核心期刊。

2.4　情报语言学著作分析

孙勇和王秀玲曾对1949－1995年我国情报语言学著作进行了统计分析,共143种,其中编著104种(占72.7%),译著39种(27.3%)。1995年以后,也陆续有著作出版,因此至今,著作总数应在150种以上。

表2－5、6、7均引自孙勇和王秀玲的研究论文,从中可以看出我国情报语言学著作的基本情况。

表2－5　我国情报检索语言著作的递增状况

年　份	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
当年种类	0	1	0	1	1	0	3	0	5	3	3	2
累计种类	0	1	1	2	3	3	6	6	11	14	17	19
年　份	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
当年种类	1	1	2	1	0	0	1	0	1	0	0	0
累计种类	20	21	23	24	24	24	25	25	26	26	26	26
年　份	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
当年种类	0	0	1	0	3	1	2	5	15	8	8	8
累计种类	26	26	27	27	30	31	33	38	53	61	69	77
年　份	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
当年种类	17	7	5	3	7	6	6	6	5	3	1	
累计种类	94	101	106	109	116	122	128	134	139	142	143	

资料来源:孙勇,王秀玲:我国情报检索语言著作研究,《中国图书馆学报》,1997(5)

表2－6　我国情报检索语言著作的内容分布

类　　别		第1阶段	第2阶段	第3阶段	第4阶段	合　　计	
		(49－54)	(55－65)	(66－76)	(77－95)	种数	比例(%)
1	分类检索语言著作	2	20	3	70	95	60.4
2	主题检索语言著作	1	1		34	36	25.2
3	分类主题一体化检索语言著作				2	2	1.4
4	综合性检索语言著作				10	10	7.0
合　计	种　　数	3	21	3	116	143	100.0
	比例(%)	2.1	14.7	2.1	81.1	100.0	
年　均　种　数		0.5	1.9	0.3	6.1	3.0	

资料来源:同表2－5

表2-7 11种情报检索语言著作的被引状况

	著作名称	面世年数	被引总数	年均被引次数
1	情报检索语言	13	88	6.8
2	图书分类	13	57	4.4
3	情报检索词汇规范化	14	59	4.2
4	主题法导论	5	21	4.2
5	汉语主题词表标引手册	11	46	4.2
6	中国图书馆图书分类法中国图书资料分类法第三版使用手册	5	20	4.0
7	主题法的理论与标引	11	40	3.6
8	情报检索与主题词表	8	28	3.5
9	现代西方主要图书分类法评述	16	54	3.4
10	图书分类学	15	45	3.0
11	中国分类主题词表教程	2	5	2.5
	合计·平均	113	163	4.1

资料来源:同表2-5

根据情报语言学著作的影响性,以及对情报语言学学科建设的影响,以下10部著作是情报语言学的经典性的核心著作:

(1)《情报语言学基础》,张琪玉编著。最早于1983年由武汉大学以《情报检索语言》为名出版,全书共8章,全面论述了情报检索语言理论与实践问题。1987年该书增订出版,增加了情报语言学一般问题,情报检索语言基本方法,情报检索计算机化与情报语言学的发展3章和文献主题的构成因素及层次一节。1997年该书增订二版出版,补充了有关分类法主题法一体化和自然语言检索两章。这是我国情报语言学最具代表性的著作,也是情报语言学的奠基性著作。据作者统计,该书至少被180名著者的约250种文献引用。

(2)《现代西方主要图书分类法评述》,刘国钧著。吉林人民出版社1980年出版。曾在1978-1980年的《社会科学战线》杂志上连载。该书是刘国钧在60年代根据教学需要撰写的,系统介绍了DDC、EC、LCC、UDC、BC、SC、CC等七部著名的分类法,对我们了解和学习国外分类理论和技术起到了积极的作用。

(3)《图书分类学》,白国应编著。书目文献出版社1981年出版,全书共17章,是作者从事图书分类研究、教学和图书分类工作的经验总结。该书资料丰富,论述详尽,是学习图书分类学的一部很好的教科书。

(4)《图书分类》,李严、史永元、张涵、侯汉清、马张华编著。书目文献出版社1983年第一版,1990年修订版。全书对图书分类法的理论、图书分类工作实践进行了全面论述,是我国使用面最广的一部图书分类教材。

(5)《图书分类学》,周继良主编,周绍萍、俞君立、张燕飞编著。武汉大学出版社1989年第一版,1992年修订版。本书建立的图书分类学体系比较完整,是一部优秀的大学教材。

(6)《图书馆分类工作手册》,侯汉清、王荣授主编,陈树年、戴维民副主编,中国科学技术出版社1992出版。全书分术语篇、工具篇、文献篇、人物篇、团体机构与会议篇、实践篇、附录等,共55万字,由全国77位专家、学者共同撰稿完成,是我国有关图书分类学最系统、规模最大的工具书。

(7)《汉语主题词表标引手册》,钱起霖主编。科学技术文献出版社1985年出版。这是针对《汉

语主题词表》的使用，专门编写的标引手册，实用性很强，它对我国以后制定的叙词标引国家标准有极大的影响。

(8)《主题法的理论与标引》，刘湘生编著。书目文献出版社 1985 年出版。这是作者致力于主题词表理论与实践研究的结晶，是我国比较早地系统阐述主题法的著作，具有开创性。

(9)《情报检索与主题词表》，丘峰著。书目文献出版社 1988 年出版，全书根据国内外 30 多种主题词表的材料，对主题法的理论、结构及编制技术等进行了阐述，其中穿插的 200 多种图表起到了补充说明作用。该书是一部颇具参考价值的著作。

(10)《主题法导论》，侯汉清、马张华主编。北京大学出版社 1991 年出版。本书 9 章，全面论述了各种类型的主题检索语言，重点介绍了叙词法的理论以及叙词标引技术。并首次将分类法主题法一体化词表作为一章写入本书，这既是一本有关主题法的专著，也是一部有关主题法的较好的教科书。

在情报语言学著作中，还出版了一批个人文集和综合性文集。正式出版的个人文集代表性的有《刘国钧图书馆学论文选集》、《杜定友图书馆出版学论文选集》、《张琪玉情报语言学文集》、《当代分类法主题法索引法研究》(侯汉清)、《情报检索语言综论》(戴维民)、《联机环境中的情报检索语言》(曾蕾)、《阮冈纳赞理论及其研究》(叶千军)等。综合性文集主要有：《情报检索语言论文选》、《图书分类教学参考资料》、《图书分类学文集》、《文献分类学论文荟萃》等。

2.5 对国外情报语言学著作翻译

我国在情报语言学研究中，非常注意吸收国外的先进理论和技术，其中国外一些著名的学者及其著作都被介绍到中国来，极大地丰富了我国情报语言学理论体系。像阮冈纳赞、杜威、兰开斯特、艾奇逊、米尔斯蒂等国外学者都成为了我国学术界熟知的学者。

国外情报语言学系统著作在我国翻译的约在 40 种以上，其中主要的有：

(1)《图书怎样分类》，刘国钧著。北京开明书店 1953 年出版。本书虽为著，其实是根据国外的有关著作编译而成。

(2)《情报检索理论概述》，(苏)切尔内著，赵宗仁、许恒泰译。科学技术文献出版社 1980 年出版。

(3)《分类目录主题索引编制法》，肖自力、李修宇、侯汉清编译，刘国钧校。书目文献出版社 1980 年出版。

(4)《情报检索词汇控制》，(美)兰开斯特著，杨劲夫等译。科学技术文献出版社 1980 年第一版。第二版由侯汉清、戴维民、陆宝树翻译，同济大学出版社 1992 年出版。

(5)《图书分类理论》，(英国)布坎南著，于鸣镝、冯频译。辽宁省图书馆学会 1981 年出版。

(6)《冒号分类法理论与实践》，(印)塞克德瓦著，吴人珊译。书目文献出版社 1985 年出版。

(7)《冒号分类法解说及类表》，宋克强、许培基译著。书目文献出版社，1986 年出版。

(8)《需要以分面分类作为一切情报检索方法的基础》，英国伦敦分类法小组著，侯汉清、张欣毅译。载《宁夏图书馆通讯》1985 年第 2 期。

(9)《分面分类法——专业分类表的编制和使用指南》，(英)维克里著，侯汉清、张欣毅译。载《宁夏图书馆通讯》1984 及 1985 年。

(10)《主题检索系统设计选择》，(美)米尔斯蒂著，叶千军、吴人珊译。书目文献出版社 1991 年出版。

(11)《国外情报检索语言研究——分类主题一体化的理论与实践》，吕其苏等编译。社会科学

文献出版社 1989 年出版。

(12)《情报组织与利用》,(美)索尔格尔著,谢新洲、陈睿、常彤译。书目文献出版社 1992 年出版。

(13)《文献标引手册及实例》,谭重安等编译,钱起霖、郭后光审校。科学技术文献出版社 1983 年出版。

(14)《图书资料的分类》,(日)宫坂逸郎、石山洋等编,宋益民译。书目文献出版社 1982 年出版。

(15)《图书分类法史略》(第 2 卷),(苏)沙姆林著,何善祥、郑盛畴译,王锡文校。科学技术文献出版社 1989 年出版。

3 人物篇

在 50 年的发展中,我国情报语言学形成了一支数量庞大,有研究核心的研究者队伍,这是学科发展和建设的重要因素。

3.1 关于核心作者队伍的研究概述

关于我国情报语言学的核心作者在 80 年中期就有专门研究,白国应、张琪玉和丘峰、翟凤岐和张芝兰、吴秀爽和赵良英、孙勇和王秀玲等都作过统计分析。

张琪玉和丘峰通过发文数量和著作被引次数确立了 1949 - 1989 年我国情报语言学研究者核心队伍,如表 3 - 1 所示:

表 3-1 情报语言学的核心作者

发文最多的作者(依次排序)	著作被引次数著者(依次排序)
白国应 杜定友 皮高品 刘国钧 李锡初 钱亚新 张琪玉 丘 峰 侯汉清 戴维民 刘湘生 叶千军 傅兰生 卢中岳 曾 蕾 李 严 史永元	张琪玉 刘国钧 白国应 侯汉清 刘湘生 丘 峰 史永元 傅兰生 曾 蕾 杜定友 钱起霖 周继良 戴维民

资料来源:张琪玉,丘峰:我国情报检索语言四十年文献综述,见《情报检索语言论文选》(书目文献出版社,1990)

孙勇和王秀玲对 1949—1995 年的情报语言学著作进行了研究,得出了我国情报语言学领域的多产作者及合著率和合作度,如表 3 - 2 和表 3 - 3 所示:

表 3-2 我国情报语言学领域的多产作者

姓 名		著作种数	代表性著作	姓 名		著作种数	代表性著作
1	侯汉清	10	主题法导论	5	陈树年	3	中国图书馆图书分类法的理论与使用
2	张琪玉	6	情报检索语言	6	刘湘生	3	主题法的理论与标引
3	白国应	5	图书分类学	7	丘 峰	3	情报检索与主题词表
4	戴维民	4	情报检索语言综论				

资料来源:孙勇、王秀玲:我国情报检索语言著作研究,《中国图书馆学报》,1997(5)

表 3-3　我国情报语言学著作的合著率与合作度

项目		作者类型								合著率(%)	合作度
		个人作者							团体作者		
		1人	2人	3人	4人	5人	6人、6人以上	小计			
1	著作种数	53	14	18	10	9	26	130	13	62.9	2.53
2	所占比例(%)	37.1	9.8	12.6	7.0	6.3	18.2	90.9	9.1		

资料来源:同表 3-2

白国应在多篇文章中提到了我国分类学的核心研究队伍,在 1995 年发表的《走向 21 世纪的中国文献分类学》一文中列出了建国后我国分类学的著名作者:刘国钧、杜定友、皮高品、钱亚新、张琪玉、白国应、张德芳、史永元、周继良、袁涌进、李兴辉、韩承铎、许培基、刘湘生、侯汉清、卢子博、曹殿举、刘延章、丘峰、李锡初、宋传山、戴维民等。

吴秀爽和赵良英著文对 1993 年之前的 10 年中主题法研究论文进行了调查和分析,根据发文量得出了一个 15 人的核心著者,如表 3-4 所示:

表 3-4　发文 5 篇以上的 15 名著者及其发文量(篇)

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
姓名	刘湘生	侯汉清	戴维民	傅兰生	王金夫	叶千军	张志升	王正兴	郭俊	陶钰	陈守福	曾蕾	程桂英	帅兵	郭友仁
发文量	15	15	13	12	9	8	8	7	7	7	6	6	6	5	5

资料来源:吴秀爽、赵良英:近 10 年我国主题法研究论文的调查与分析载,《图书情报工作》,1993(1)

翟凤岐和张芝兰对 1949-1991 年我国主题检索语言进行系统分析和研究,得出我国主题检索语言论文作者知名度较高的有 7 人,如表 3-5 所示:

表 3-5　我国主题检索语言论文作者知名度较高的 7 人

姓名		发文量最多的 5 人	篇均引文量最高的 5 人	篇均被引量最高的 5 人
1	戴维民	(18 篇)	(6.1 条/篇)	
2	侯汉清	(29 篇)	(4.0 条/篇)	(2.8 条/篇)
3	刘湘生	(29 篇)		(3.0 条/篇)
4	丘　峰	(35 篇)	(3.6 条/篇)	(1.8 条/篇)
5	叶千军		(1.9 条/篇)	
6	曾　蕾		(2.2 条/篇)	(1.7 条/篇)
7	张琪玉	(16 篇)		(8.7 条/篇)

资料来源:翟凤岐、张芝兰:我国主题检索语言的理论与实践(1949-1991)概述,《中国图书馆学报》,1993(2)

3.2　关于核心作者队伍的最新统计分析

根据张琪玉《情报语言学文献库》的作者统计分析,50 年来情报语言学的个人作者总数为 4516 人,其中发文一篇的作者数为 3153 人,发文 2 篇以上共 1363 人,发文 10 篇以上的 98 人,发文 20 篇以上的 28 人,如表 3-6,表 3-7 所示。

表 3－6　发文 20 篇以上的作者

排名	作者	发文量	排名	作者	发文量
1	张琪玉	146	15	丘　峰	31
2	侯汉清	136	16	陈晓华	30
3	白国应	88	16	李　严	30
4	戴维民	85	18	李锡初	28
5	皮高品	62	19	刘延章	26
6	刘国钧	51	20	高崇谦	25
7	杜定友	50	21	曹树金	24
8	钱亚新	48	21	吴声亮	24
9	陈光祚	44	21	俞君立	24
10	刘湘生	38	21	曾　蕾	24
10	叶千军	38	21	王永成	24
12	傅兰生	32	26	赖茂生	23
12	蔡振华	32	27	陈树年	20
12	史永元	32	27	黄恩祝	20

资料来源：张琪玉：《情报语言学文献库》(电子版)，武汉大学出版社，1998

表 3－7　发文 10－18 篇的作者

序号	作者	作者数	发文量
1	赵宗仁　周智佑	2	18
2	洪　漪　顾耀芳　林德海　李兴辉 卢子博　钱　唐　王金夫　张正强	8	17
3	黄祥喜　曾民族　肖自力　赵燕群	4	16
4	陆近春　李景正　钱　亮　马张华 王　源　姚行地　张燕飞	7	15
5	陆宗城　王知津　王正兴　张　进 谢新洲　章元峰　朱孟杰	7	14
6	关家麟　贺广明　陈守福　黄景行 卢中岳　秦　明　苗绍武　倪　波 杨则正　杨晓骏　张志升　赵怀生 周全明	13	13
7	达　翰　郭　俊　黄　钢　宋明亮 汪东波　孙绍荣　王　兵　邵品洪 陶　钰	9	12
8	程　磊　毕　强　程长源　陆长旭 黄　明　黄俊贵　刘植惠　王克明 田文清　曾世荣　武宁生　周继良 鄢德梅	13	11
9	崔明爱　曹　之　邓汉成　程桂英 邓福泉　柳晓春　刘洪波	7	10
共计		70	

资料来源：同表 3－6

根据表 3－6，表 3－7 所示，可以初步确定发文在 20 篇以上的 28 名作者为我国情报语言学的第一核心作者群，对情报语言学学科建设具有重要贡献。发文 10 篇以上，20 篇以下的 70 名作者，

是情报语言学的第二核心作者群,第二核心作者群核心数量较第一核心作者群数量大,是推动情报语言学学科发展的中坚力量。

在情报语言学文献中,有一些是集体责任者,根据统计频率,主要的如表3-8所示:

表3-8 情报语言学主要集体责任者

序号	集体责任者	责任频率
1	中国图书馆图书分类法编辑组	54
2	中国图书馆图书分类法编辑委员会	54
3	图书分类法编辑小组	45
4	中国科学技术情报研究所	18
5	北京图书馆	15
6	武汉大学图书馆学系	12
7	中国专利局审查部	9
8	全国文献工作标准化技术委员会	8
9	中国人民大学图书馆	8
10	全国文献工作标准化技术委员会第五分会	7
11	第三机械工业部六二八研究所	7
12	中国科学院图书馆	7
13	七四八工程情报检索办公室	6
14	冶金工业部情报研究所	6
15	南京图书馆	5
16	北京图书馆分类法词表组	5

资料来源:同表3-6

3.3 我国情报语言学领域最具影响的作者

上文指出的情报语言学的第一核心作者群是情报语言学学科的主要建设者,其中有些堪称情报语言学的奠基者和开拓者,是情报语言学学术带头人。在《中国大百科全书》图书馆学情报学档案学卷所载当代学者条目中,其中有5位主要贡献在情报检索语言方面,他们是:刘国钧、杜定友、皮高品、钱亚新和张琪玉。

(1)刘国钧(1899-1980) 我国现代著名的图书馆学家,我国现代图书馆事业奠基人之一。在情报检索语言方面,早在1929年就编制了《中国图书分类法》,建国后,这部分类又经修订,并被北京图书馆使用。经赖永祥等人多次修订在台湾仍是使用最广泛的分类法之一。建国后,刘国钧致力于新型分类法编制工作,包括《中小型图书馆图书分类表草案》、《中国图书馆图书分类法(大型法)》等,同时对分类理论,分类技术进行了深入研究,对指导我国分类实践起到了重要作用。特别是研究和介绍国外分类法及其理论开阔了我们的眼界,对缩短我们与国外的差距非常有益。

(2)杜定友(1898-1967) 中国现代著名的图书馆学家,图书馆学教育家。图书分类法是杜定友研究的主要方向之一,早在1936年就编成了《杜氏图书分类法》,后改名为《世界图书分类法》,具有很大影响。建国后杜定友积极参与新分类法的编制,同时进行理论研究,发表了许多有价值的研究成果,如《新图书分类法刍议》、《新图书分类法草案》,其中提出以毛泽东关于知识分类理论,建立哲学、自然科学、社会科学三大部类体系,对以后我国图书分类法的编制产生了很大影响。杜定友在分类法方面有很多创新思想,其中提出了关于分类法与主题法结合的问题,提出了组配分类法的思想,这在他编制的《图书资料分类法》中得到充分体现,这部分类法最终虽未能完成,但杜定友在分类法方面的创新思想显而易见。

(3)皮高品(1900－1998)　中国现代著名的图书馆学家。长期从事图书分类学的研究和教学工作。他1934年编制出版的中英对照《中国十进分类法及索引》是一部根据中国学术特点而编，结构完整，便于类分中外图书的分类法。建国后参与分类法的编制工作，主张按照恩格斯有关物质运动形态发展规律的学说，建立科学的图书分类法，其基本序列应为哲学、社会科学、自然科学和综合性图书。皮高品发表了大量图书分类法评论性文章，提出了许多独到的见解，他的评论文章汇编成《图书分类法评论选集》。

(4)钱亚新(1903－1990)　中国现代著名图书馆学家、目录学家。钱亚新一向重视对图书分类学的研究，早年就积极介绍国外的分类法，如《杜威法》、《美国国会图书馆图书分类法》、《布鲁塞尔分类法》、《冒号分类法》等。直到晚年，仍在关心我国图书分类法的研究和修订工作，写了许多论文，提出了在《中图法》中建立"综合性学科类"的建议。钱亚新在索引法和汉字排检法及著者号码表方面也进行了卓有成效的研究。

(5)张琪玉(1930－)　中国当代著名图书馆学家，情报语言学的主要开拓者之一。从1979年开始，张琪玉集中精力，致力于情报检索语言研究，写成了《情报检索语言》一书，并在武汉大学成功地开设了情报检索语言课程。他认为情报检索语言是为在情报检索中求得较好的检索效果而创制的人工语言，它作为情报检索系统的一个组成要素，在情报检索中起语言保证作用。由此，他把各种类型的情报检索语言进行综合分析，抽象升华，横向比较，把研究的基点建立在提高情报检索效率上。从而构建了情报语言学学科的基本框架，研究方法体系，拓展了情报语言学新的研究领域，不断进行学科的创新和新型情报检索语言的创制研究。

4　分类法篇

早在西汉，《七略》就建立了图书分类的七分之法，到唐朝《隋书经籍志》又建立了经、史、子、集四分法，而清朝的《四库全书总目提要》把我国四分法体系建立得相当完整，至今仍为古籍书目之实用体系。

而至近代，由于西学东渐，特别是《杜威十进分类法》的传入，我国一大批学者，如沈祖荣、洪有丰、杜定友、刘国钧、皮高品、何日章、裘开明和王云五等人编制了一批"仿杜"、"补杜"和"改杜"分类法，形成了我国近代分类法的主要特色，虽有照搬的明显痕迹，但在技术上为我国建国后编制新型分类法奠定了一定的基础。

建国后，由于社会形态的根本变革，使我国分类法的编制、使用及理论研究都翻开了崭新的一页。

4.1　分类法的编制和修订

建国初期，我国就开始了新分类法的编制，同时对旧分类法进行修订以应急需。

1948年8月东北图书馆(今辽宁省图书馆)所编的《图书分类法》(简称《东北法》)正式出版，1951年出修订本。这是建国后编制的第一部分类法，第一次将毛泽东著作和鲁迅著作辟为特藏类目。1950年10月油印出版，1951年正式出版的《图书分类法》(简称《山东法》)在一些类目下采用了按观点区分的方法。《东北法》和《山东法》在建国之初由于编修及时，从而在我国当代分类法史上占有重要地位。

50年代掀起了我国分类法编制的一个高潮，编制的分类法总共达30部之多，基本满足了图书分类工作的需要，这期间编制的有影响的分类法有：《广东人民图书馆图书分类法》、《中国人民大学

图书馆图书分类法》、《中小型图书馆图书分类简表》、《浙江图书馆图书分类新表》、《中国科学院图书馆图书分类法》、《武汉大学图书分类法》等。

50年代还对有些旧分类法进行修订后继续使用，如金云铭1928年编制的《中国图书分类法》于1955年修订，刘国钧1929年编制的《中国图书分类法》由北京图书馆于1957年修订，并在北京图书馆使用。

由于分类法编制与使用各行其是，对统一使用分类法极为不利。从60年代开始，图书馆界极力寻求解决统一分类法问题，1960年编制的《大型图书馆图书分类法草案》就是一个重要努力，虽然编制工作半途而废，但为70年代《中国图书馆图书分类法》的编制奠定了基础。1975年《中图法》第一版的出版，开始了我国统一分类法的新时代。80年代以后，我国分类法实际上已基本走向统一，从分类法的编制转到及时修订分类法上来。但其间编制的《标准文献分类法》和《中国档案分类法》等是针对不同文献类型的重要分类法。

我国建国后编制的主要分类法详见表4-1。

表4-1 建国后编制的主要图书分类法

序号	图书分类法名称	编辑者	首版年	备注
1	东北图书馆图书分类法	东北图书馆	1949	后改辽宁省图书馆
2	山东图书馆图书分类表	山东图书馆	1950	
3	苏北图书馆图书分类表	苏北图书馆	1950	后改名为苏州市图书馆
4	新图书分类法	杜定友	1950	
5	新时代应用图书分类法	卞宝弟	1950	
6	农业科学图书分类法	陈士宗	1950	
7	广东人民图书馆图书分类法	广东人民图书馆	1952	后改广东省中山图书馆
8	中国人民大学图书馆图书分类法	中国人民大学图书馆	1952	
9	中文图书分类法	南京图书馆	1954	
10	图书分类法	西北农学院图书馆 范世伟	1954	
11	图书分类表	金天游	1954	
12	中国图书分类法	金云铭	1955	修订版，1928年编成
13	浙江图书馆图书分类新表	浙江图书馆	1956	
14	中国图书分类法	刘国钧原编 北京图书馆修订	1957	修订版，1929年编成
15	中小型图书馆图书分类表草案	中小型图书馆分类表 编辑小组	1957	
16	中国科学院图书馆图书分类法	中国科学院图书馆	1958	
17	大型图书馆图书分类法	陕西省图书馆蒋一前	1958	根据《中小型表》扩充
18	武汉大学图书分类法	武汉大学图书馆学系	1959	
19	上海社会科学院图书分类法：草案	上海社会科学院图书馆	1960	
20	图书分类表	吉林省图书馆	1960	
21	大型图书馆图书分类法 又名：《中国图书馆图书分类法》(草案)	大型图书馆图书分类法 编委会	1960	1960年完成初稿，1963年5月改名中国图书馆图书分类法草案
22	中国图书馆图书分类法及系列版本	中国图书馆图书分类法 编委员会	1975	1999年第四版改为中国图书馆分类法
23	中国图书资料分类法	中国图书资料分类法编辑组	1975	在《中图法》的基础上加细和修订而成。
24	上海戏剧学院专用图书分类法	上海戏剧学院图书馆	1980	

序号	图书分类法名称	编辑者	首版年	备注
25	法学资料分类法	王金祥	1984	
26	社会科学资料分类法	中国社科院经济研究所	1985	
27	军事图书分类法	军事科学院图书馆 杨　健	1986	
28	中国标准文献分类法	国家技术监督局	1989	
29	中国新闻资料分类法	该法研究小组	1992	
30	中国档案分类法	该法编委会	1997	

资料来源:[1]白国应:中国图书分类学四十年,《高校图书情报学刊》,1989(1)
[2]北京图书馆图书馆学研究部:《全国中文图书馆学情报学档案学书刊资料联合目录》(草稿),1987
[3]有关分类法

我国50年分类法编制、修订的主要经验是:(1)坚持以马列主义、毛泽东思想、邓小平理论为指导思想。具体地表现为,在分类法类目体系中突出马列主义、毛泽东思想类目,几部有代表性的分类法都首先序列了"马克思列宁主义、毛泽东思想"大类,《中图法》第四版还在这一大类中增列了"邓小平理论"有关类目;恰当地运用观点区分;在整个宏观体系中,遵循毛泽东关于知识分类的论述,即分为哲学、社会科学、自然科学,并在此基础上形成了我国图书分类法的"五分法"体系。(2)辩证地处理了分类法的"思想性"、"实用性""科学性"的三性原则。做到既体现分类法的思想性,又科学合理地组织类目,以符合我国图书分类的实际需要,80年代以后又提出了"网罗性"、"专指性"、"稳定性"、"文献保证和用户保证"等原则方法。(3)集中全国力量,集体编制。从而保证了分类法编制的质量,保证了分类法的维护和管理正常化,这也有利于统一分类法。(4)积极地吸取和总结国外的先进经验。建国初期全面学习了前苏联的分类法编制技术,对建立我国分类法体系起到了积极的作用。70年代末以来对英美及其他国家分类法理论与技术的借鉴也大大提高了我国分类法的编制水平。

4.2 几部重要的分类法

1.《中国人民大学图书馆图书分类法》

简称《人大法》。张照、程德清主编,中国人民大学图书馆集体编著。1951年起开始编制,1953年出版第一版,以后又多次修订,至1996年已修订出版第六版。各版情况如表4-2所示。

表4-2 《人大法》各版情况

版次	主编	出版年月	开本	页数	印刷部数
初稿	张　照	1952年10月	32	182	400
第一版	张　照	1953年9月	32	144	
第二版(初稿)	张　照	1954年6月	32	258	4370
第二版(增订本)	张　照	1955年8月	32	270	13289
第三版	张　照	1957年7月	32	736	31077
第四版	张　照	1962年4月	32	1294	
第五版	张　照 程德清	1982年2月	32	720	13000
第五版 社会科学	张　照 程德清	1989年10月	32	390	3000
第六版	杨东梁 冯之圣	1996年3月	16	354	2000

资料来源:白国应:《中国人民大学图书馆图书分类法》的历史、功绩、问题和展望,《图书情报论坛》,1998(3)

白国应对《人大法》作了比较全面的评价，总结了它主要功绩和特点：(1)首先提出了社会主义图书分类法编制的思想性、科学性和实用性原则；(2)首次应用马克思主义、列宁主义、毛泽东思想关于科学分类的理论组织分类体系，确立四大部类，十七大类的分类框架；(3)首先将"马克思列宁主义、毛泽东著作"单独作为基本大类，并序列为十七大类之首；(4)创立了新的号码制度——展开层累制；(5)哲学社会科学类目比较丰富、新颖、具有特色；(6)为类目作详细的注解，开创注释体例的新局面；(7)结构比较完整，有出版说明、大纲、简表、基本类目表、详表、复分表、索引、附录等；(8)为建国初期新成立的图书馆所采用，解决了燃眉之急；(9)参加了《计算机文献标引对照系统》，实现了《人大法》与《中图法》、《科图法》的相互对照和转换，为建立全国文献信息网准备了条件；(10)学术影响比较深远。

2.《中小型图书馆图书分类表草案》

简称《中小型表》。1956年开始编制，于1957年由文化部社会文化事业管理局公布使用，并迅速为全国中小型图书馆广泛采用。至1965年又多次修订充实。该法主要供藏书在10万册以内的中小型图书馆使用。并设想《中小型表》"予以缩小，就可以成为小型图书馆的图书分类法，予以扩大，就可以成为大型图书馆的图书分类法，从面为编制统一的图书分类法打下基础"。这部分类法编辑小组的成员刘国钧、杜定友、张遵俭、钱亚新、袁涌进和李兴辉都是老一辈著名的分类法专家，代表了当时分类法编制的最高水平。编者之一李兴辉先生总结这部分类法具有四个特点：(1)以马克思主义关于科学分类的理论为指导思想，首创以"五分法"为编制基础，为我国其后编制的分类法所仿效，在马克思列宁主义、哲学、社会科学、自然科学、综合性图书五个基本部类的基础上再展开为21个大类，进面展开三、四级类目；(2)重视科学分类原则，分类概括而明确，逻辑性较强。如学科的理论和观点区分统一用总论复分方法处理，国家地区的序列以自然区划分为依据；(3)在我国首先采用字母与数字混合号码的层累标记制度，用字母标记大类，不仅适应了分类法类目展开的需要，同时加强了分类法的助记性。在数字部分基本采用层累制，尽可能使号码的级位表达类目的级位，但不受层累制的约束，适当地运用了借号法，号码简短、易记、易用；(4)类目注释比较详细，对类目的内容范围和分类方法作了必要的说明，说明文字简练、清晰，适宜中小型图书馆使用。

3.《中国科学院图书馆图书分类法》

简称《科图法》。中国科学院图书馆编制，科学出版社出版，是一部大型综合性分类法。1954年开始编制，1958年正式出版，现已出第三版。各版情况如表4－3所示：

表4－3 《科图法》各版情况

版　次	出版者	出版年月	页数	开本
自然科学部分初稿	油印本	1957年	392	16
第一版	科学出版社	1958年11月	258	16
索引	科学出版社	1959年10月	708	16
自然科学、综合性图书部分第二版	科学出版社	1974年2月	420	16
哲学、社会科学部分第二版	科学出版社	1977年11月	234	16
第三版	科学出版社	1994年12月	844	16

《科图法》设有马克思列宁主义毛泽东思想、哲学、社会科学、自然科学、综合性图书五大部类，

共分 25 大类。标记采用单纯阿拉伯数字，分为两部分：第一部分采用顺序制，用 00－99 标记五大部类 25 个基本大类及其主要类目，第二部分采用“小数制”，即在 00－99 两位数字后加一小数点“·”。小数点后基本按小数制编号。

《科图法》第二版国内曾有 1000 多个单位使用，主要是中国科学院系统的图书馆。该法受到用户和学术界的充分肯定和好评，认为这部分类法自然科学类目比较详细，科学系统性比较强，能较好地反映当时科学技术的发展水平，标记制度单纯简洁，编制体例清楚，结构比较完整，能够类分中外、古今图书，实用性比较好。因此被公认是我国三大图书分类法之一。该法 1986 年获国家科委颁发的科技情报成果二等奖和中国科学技术进步三等奖。

为了使《科图法》更加科学实用，成立了三版修订委员会，从 1987 年开始修订工作，修订后的第三版于 1994 年出版。对第三版的修订主要在九个方面：(1)改变部分类目的体系；(2)增加新类目；(3)修订原类目；(4)统一编制体例；(5)修订补充附表；(6)修订补充专类附表；(7)增加大类说明；(8)修改类目注释；(9)正确处理思想性的一些问题。修订后的第三版更加科学合理，是一部优秀的大型综合性类法。

4.《大型图书馆图书分类法》

又名《中国图书馆图书分类法》(草案)，简称《大型法》。1959 年在中央文化部和教育部的领导下，组成编委会，开始编制工作，至 1963 年 5 月完成了下册自然科学部分，1966 年 3 月完成了上册哲学、社会科学部分。由于文革动乱的原因，最终未能正式完成，但在编制《大型法》过程中对大型综合性分类法相关问题进行了深入的讨论和研究，其所采用的五大部类进面展开的 21 个基本大类，以及字母和阿拉伯数字相结合的混合号码及其相关标记技术基本上为后来编制的《中图法》所继承。因此，它在我国当代图书分类法史上具有一定的影响和地位。

李兴辉概括其有三个特点：(1)广泛组织了全国各系统图书馆，科研单位和教学单位的专家学者参与研究和编制。在编制过程中，对分类法的体系结构、标记符号以及编制技术，展开了多方面的探讨，为这部分类法的编制创造了良好条件。(2)类目较详细，哲学、社会科学类目伸展到五、六级，自然科学类目伸展到七、八级，所收的类目达 65000 余条，适应了科学技术发展在文献分类上的需要，同时也解决了我国古籍分类的问题。(3)吸取了中外图书分类法的编制经验，如：编制交替类目、采用组配编号法，重视复分表的编制和类目之间的仿照复分等，提高了分类法的适应性和助记性，增强了标引文献的能力，方便专业图书馆和情报单位使用。

5.《中国档案分类法》

简称《中档法》。编制工作开始于 1983 年，1986 年完成编制，1987 年出版第一版。《中档法》第一版实际上是一个简本性质的试用本，一开始就决定经试用后准备修订出版大型的《中档法》。

修订工作开始于 1989 年，近千人参加，在 53 个专业表的基础上编制而成，类目数达 10 万条。1995 年通过鉴定，并于 1997 年 12 月正式出版《中档法》第二版。

《中档法》“一法四表”，即由四个分类表组成，分别是《中华人民共和国档案分类表》、《新民主主义档案分类表》、《民国档案分类法》和《清代档案分类表》。每个表都由主表和辅助表组成。各表的基本大类设置和辅助表设置如表 4－4 所示。

表 4-4 《中档法》各表基本大类及辅助表

	中华人民共和国档案分类表	新民主主义分类表	民国档案分类表	清代档案分类表
基本大类	A 中国共产党党务 B 国家政务总类 C 政法 D 军事 E 外交 F 政协、民主党派、群众团体 G 文化、教育、卫生、体育 H 科学研究 J 计划、经济管理 K 财政、金融 L 贸易、旅游 M 农业、林业、水利 N 工业 P 交通 Q 邮电 R 城乡建设、建筑业 S 环境保护、土地管理 T 海洋、气象、地震、测绘 V 标准、计量、专利	A 中国共产党党务 B 政务 C 公安、司法、法院 D 军事 E 国际共运、外事 F 群众团体、群众运动 G 文化、教育、卫生 J 经济管理 K 财政、金融 L 商业 M 农业 N 工业 P 交通、邮电	A 中国国民党党务 B 政务总类 C 内政 D 社会 E 考试、监察 F 司法 G 军事 H 外交 J 教育、文化、科学研究 K 经济总类 L 财政、金融 M 农、林、水利 N 工、矿、电业 P 商业 Q 交通 R 邮电	A 政务总类 B 宫廷、皇族及八旗事务 C 职官、吏役 D 军事 E 政法 F 民族事务 G 中外关系 H 镇压人民斗争活动 J 宗教事务 K 文化、教育、卫生、科学研究 M 财政 N 金融 P 农业、水利、畜牧业 Q 手工业、工业、公用事业 R 建筑 S 交通、邮电 T 商业 N 天文、地理
辅助表	综合复分表 世界各国和地区表 中国地区表 中国民族表 科技档案复分表	综合复分表	综合复分表 世界各国和地区表 民国时期行政区划表	综合复分表 世界各国和地区表 清代行政区划表 中国民族表

《中档法》第二版坚持职能分类原则，并从我国档案资源构成的实际情况，确定"一法四表"的体系结构和详略情况，通过采用多种分类法编制技术以期达到提高《中档法》的检索功能，使之成为一部性能优越的档案分类检索语言。

6.《中国图书馆分类法》

原称《中国图书馆图书分类法》，第四版改名《中国图书馆分类法》，简称《中图法》。编制于70年代，是我国一部大型综合性分类法。1971年北京图书馆等36个单位组成编辑组开始编制，1973年3月编成试用本，1975年10月科学技术文献出版社出版第一版。1979年7月成立中国图书馆图书分类法编辑委员会进行修订，1980年6月由书目文献出版社出版第二版，1990年出版第三版，1999年出版第四版。《中图法》各版情况如表4-5所示：

表 4-5 《中图法》各版情况

序号	版本	出版者	出版年	页码	开本
1	试用本	北京图书馆	1973	673	16
2	第一版	科技文献出版社	1975	611	16
3	第二版	书目文献出版社	1980	674	16
4	第三版	书目文献出版社	1990	762	16
5	第四版	北京图书馆出版社	1999	1141	16

为了使《中图法》更好地适应各类各级图书馆及各种文献类型分类的需要，编制了各种配套的版本，如表4-6所示。

表 4-6 《中图法》配套版本一览表

全 称	简 称	版 本	出版年	使用对象	出 版 者
《中国图书馆分类法》	《中图法》	第四版	1999	大型馆	北京图书馆出版社
《中国图书资料分类法》	《资料法》	第三版	1990	文献资料	科学技术文献出版社
《中国图书馆图书分类法(简本)》	《简本》	第三版	1991	中小型馆	书目文献出版社
《中国图书馆图书分类法(期刊分类表)》	《期刊表》	第一版	1993	期 刊	书目文献出版社
《中国图书馆图书分类法(儿童图书馆中小学图书馆版)》	《中图法儿童中小学馆版》	第一版	1991	儿童馆、中小学馆	
《中国分类主题词表》		第一版	1994		华艺出版社
《中国图书馆图书分类法》(第二版)索引	《中图法索引》	第一版	1984	中图法资料法用户	书目文献出版社
《中国图书馆图书分类法》《中国图书资料分类法》第三版使用手册		第一版	1991		书目文献出版社

除上述不同版本和配套工具外,一些单位和部门还编制了《中图法》系列专业类表,如《地质专业分类法》、《教育专业分类法》、《公安科学分类法》等。这些专业表与《中图法》是完全兼容的,能较好地适应某种专业文献的分类标引需要。

《中图法》是我国当代分类法中编制得最为优秀者,它充分吸取了我国建国后编制大型综合性分类法的经验,全面地运用了现代分类法技术,加上它的有效管理和及时修订,不同版本和系列产品建设,而成为受用户广泛欢迎的分类法,事实上成为我国标准的统一的分类法。它所开创的我国分类法统一应用的局面可以称之为《中图法》的新时代。

4.3 对国外分类法的翻译和研究

近现代分类法起源于美国和西方国家,至50年代以后,当时苏联、印度以及日本分类法理论研究和编制技术都相对比较先进,因此引进和借鉴国外的先进理论和技术对我国分类法的理论研究和新型分类法的编制具有直接意义。

建国之初,主要是向当时的苏联学习,翻译了一些苏联的分类著作和几部分类法。安巴祖勉,托罗帕夫斯基、杰斯林科都成为我国比较熟知的专家学者。《苏联图书十进分类法》(改订本)、《十进分类法》等被翻译成中文供参考使用。建国之初,我国新型分类法的编制思想主要是借鉴前苏联的做法。

建国后,系统全面地研究国外分类法应推刘国钧教授,他在教学和指导研究生的过程对英美主要分类法进行了分析研究,分别对DDC、EC、LCC、DDC、SC、CC、BC等七部分类法作了介绍,于1964年8月写成了《现代西方主要图书分类法评述》一书,由于随后的文革动乱,该书未能及时出版,直到文革后才在《社会科学战线》上陆续刊出,随后列入《社会科学战线丛书》由吉林人民出版社出版。这部著作发表和出版之后产生了很大影响,成为经典之作。

至80年代以后,我国学术界逐步走向成熟,学术界开始比较全面的研究国外分类法理论和技术、侯汉清、吴人珊、许培基、宋克强、戴维民等人这方面的研究成果更多,影响更大一些。

前苏联沙穆林的《图书与书目分类法史略》、英国维克里的《分面分类法—专业分类表的编制和使用指南》、布坎南的《图书馆分类理论》、印度塞克德瓦的《冒号分类法理论与实成》、英国伦敦分类

法研究小组的《需要以分面分类作为一切情报检索方法的基础》等经典文献都被译成中文发表或出版，从而丰富了我国分类法理论。

4－7　中国翻译出版的国外重要图书分类法

序号	书　　名	著　者	译　者	出 版 者	出版年
1	苏联大众图书馆十进分类法	(苏)托罗帕夫斯基		文物参考资料	1950
2	苏联小型图书馆适用十进分类法	(苏)克列诺夫		人民日报社	1951
3	机械电机图书分类表	(比利时)拉芳田、奥特勒	魏 以 新	浙江省图书馆通讯	1953
4	论苏联图书分类草案	(苏)杰斯林科	范文津等	文化部社会文化事业管理局	1955
5	苏联图书十进分类法改订本	(苏)安巴祖勉等	北京图书馆编目组	北京图书馆	1956
6	十进分类法	(苏)托罗帕夫斯基	中苏友好协会总会图书资料室	时代出版社	1956
7	图书分类表——供大众图书编制标准目录和铅印卡片之用	(苏)安巴祖勉	文 敏 等	中华书局	1958
8	国际十进分类法简表	(比利时)拉芳田、奥特勒	中国科学技术情报研究所	中国科学技术情报研究所	1958
9	国际十进分类法：自然科学、应用科学		中国科学技术情报研究所	中国科学技术情报研究所	1960
10	日本专利分类表		中国科学技术情报研究所	中国科学技术情报研究所	1972
11	国际专利分类表		中国科学技术情报研究所	中国科学技术情报研究所	1972
12	国际物理学分类表	国际科学联合会文摘部	中国科学院图书馆计算机组	中国科学院图书馆	1981
13	扩展式排序体系	FID/SO 编辑组	林德海等	北京图书馆	1986
14	苏联图书馆图书分类法		倪波等译	吉林图书馆学会	1986
15	冒号分类法解说及类表		宋克强、许培基	书目文献出版社	1986
16	国际十进分类法类表		UDC 类表编译组		1989

资料来源：[1]白国应：中国图书分类学四十年，《高校图书情报学刊》，1989(1)
[2]有关分类法

4.4　分类法编制理论与技术研究

由于编制新型分类法的需要，开展了大量的理论研究，对分类法的编制技术也进行了有益的探讨，可以说涉及到分类法理论和技术的各个方面。

1950 年 6 月 2 日，文化部副部长郑振铎亲自主持召开了图书分类法座谈会，于光远、王冶秋、王重民、吕淑湘、徐特立、陈鸿舜、孙伏园、程德清、潘光旦等著名学者和图书馆学专家出席了会议。这次会议讨论五个问题：(1)要编制怎样一种性质的分类法？(2)马列主义、毛泽东思想的书籍，是按其性质分入各类，还是单独成立一个文库？(3)用十进分类法，还是新十进法，如何排列次序？(4)自然科学与应用科学分列还是并为一类？(5)宗教有无单独成一类的必要？这是我国建国后召开的有关分类法的第一次学术会议。此后召开的各种学术会议和工作会议比较多，在理论上产生

较大影响的有“铜陵会议”和“天津会议”。

1985年在安徽铜陵召开的“《中图法》(第三版)修订研讨会”就《中图法》和《资料法》中如何扩大使用组配方法和是否要为综合性科学立类问题进行了研究。会议代表们讨论认为,《中图法》和《资料法》应扩大使用组配,并提出了许多具体意见。关于综合性科学列类问题有两种意见,一种意见认为综合性科学客观存在,应列为大类或作为其他大类的类组出现,以适应分类工作需要;另一种意见认为综合性科学有些尚无定论,或在其他类中列类,因此不需要单独设立大类。《中图法》三版出版时扩大了组配方法的使用,但没有为综合性科学列类。但1996年出版的《人大法》第六版将“综合性图书”大类改为了“综合性科学、综合性图书”类组。

1991年1月在天津召开的“全国《资料法》分面改造学术研讨会”主要是为《资料法》分面改造决策提供方案及参考意见。会议对《资料法》分面改造的必要性及重要性、分面改造的路向和原则,以及技术方法等三个方面的的问题进行了研讨。大家认为《资料法》必须进行分面改造以适应发展的需要。建议可采用UDC、BC2、BBK、分面叙词表模式或集各法之长来对《资料法》进行分面改造。会议提出《资料法》应通过分面改造提高标引新主题、复杂主题文献的能力、实现多途径检索的目标等。这是一次重要的学术会议,由于《中图法》第四版修订时与《资料法》分面治之,《资料法》分面改造未能如期完成。

综合我国分类法理论和编制技术研究的成果,其研究的重点主要集中在如下方面:(1)图书分类法体系的建立,包括基本部类及其序列的确立,基本大类的划分,马克思列宁主义、毛泽东思想类目的设立,社会科学和自然科学部类的序列,按国家划分类目,依人列类,区分观点,传记类图书的集中与分散,综合性科学列类等方面的问题;(2)体系分类法的研究,包括:“集中与分散”的矛盾以及处理这个矛盾的方法,分面分析理论及体系分类法的分面改造技术,复分表、交替类目、注释、参见等的研究;(3)关于分类法思想性、实用性、科学性“三性”原则的研究,以及分类法提高情报检索效率的基本方法研究;(4)分类标记的研究,包括标记符号,标记制度,以及双位数法、八分法、借号法、字母标记法等方法技术;(5)分类法标准化问题;(6)图书分类自动化及机读分类表问题;(7)关于分类目录的研究,包括分类排架和组织分类目录职能的协调、分类目录与主题目录的关系及分类目录前途问题;(8)关于图书分类实践若干问题研究,包括各类、各科文献的分类,新学科文献分类,同类书区分,图书改编等;(9)关于基本理论研究,包括分类法的概念逻辑基础和知识分类基础,图书分类历史,图书分类学(文献分类学)等;(10)关于国外图书分类法及其理论和技术的研究等。

5 主题法篇

主题法是各种主题检索语言的一个统称,有时又专称叙词语言。

主题法最早的类型是标题法,后有单元词法,又演变为叙词法,从国外的历史看已有100多年的历史。我国主题法起步较晚,建国前后只有程长源等人涉猎过标题法,但应用非常有限。因此,本篇所述我国主题法发展情况专指叙词法的发展。在第6篇中将专门论述分类法主题法一体化词表,本篇仅概述传统模式的叙词法发展情况。

5.1 叙词表的编制与应用

早在1947—1950年期间,美国穆尔斯(Mooers,C.N.)就提出了叙词法思想,1959年美国杜邦公司编制了第一部叙词表,从60至80年代是国外叙词表编制的高峰期,进面使叙词语言成为当代情报检索语言的主流。

1964 年 5 月,我国航空部编印《航空科技资料主题表》,它是我国建国后编制的第一部主题词表,但属标题表性质。至 1971 年 12 月,该词表第二版出版时,才标志着我国叙词语言有了进一步的发展。这其中最值得一提的是《汉语主题词表》的编制。

《汉语主题词表》是作为"七四八"工程,即"汉字信息处理系统"的配套项目而编制的,于 1979 年由科技文献出版社出版,至目前为止,它仍是世界上最大规模的叙词表。由于《汉语主题词表》采用人海战术编制而成,从而锻炼了编表队伍,为我国叙词语言的发展作了人才准备。70 年代编成的叙词表还有《电子技术汉语主题表》、《原子能科技资料主题词典》、《机械工程主题词表》等约 6、7 部。

80 年代是我国叙词语言的大发展时期,编制的叙词表至少在 6、70 部以上。因此,80 年代中期,曾有一场分类目录与主题目录之争,就有人认为主题目录要取代分类目录。由于叙词语言的发展,全国汉语叙词表发展方向研讨会于 1988 年 12 月在北京召开,以期研讨叙词语言有关技术问题。会上杨南征收集了 29 部词表并进行了统计分析,其中收词 1 万以上的大型词表就有 11 部。从 80 年代后期开始,寻求编表模式和编表技术的突破,其中分类法主题法一体化和运用计算机技术编表是重点,并均取得了实用性成果。

90 年代以来,我国叙词表编制的大发展的势头不减,编制的叙词表也有几十部之多,其中代表性的有《军用主题词表》及其系列专业词表、《中国档案主题词表》等。1996 年 7 月在北京召开第二次全国分类法、叙词法发展方向研讨会,不仅展示了 90 年代的理论成果,也展示了实践成果。

到目前为止,我国已编制的叙词表在百部以上,几乎覆盖了各专业和各种文献类型,构成了我国叙词语言体系。下面表 5-1 是我国 70 年代以来各时期主要的叙词表基本情况。

表 5-1 国内几种主要叙词表

名称	编者	出版时间	专业范围	词量	
				叙词	非叙词
航空科技资料主题表	航空工业部情报系统	1971	宇宙航行;航空学	4330	199
电子技术汉语主题词表	第四机械工业部第一研究所	1977	电子工业	7500	1000
常规武器专业主题词表	常规武器系统二十个单位共同编制	1977	常规武器	4398	1625
国防科学技术主题词典	国防科委情报所	1978	国防科技	17173	3719
原子能科技资料主题词典	二机部情报所	1978	原子能科技	15179	3363
机械工程主题词表	一机部情报所	1979	机械工程、电机工程、仪器仪表	9667	1533
汉语主题词表	中国科学技术情报所 北京图书馆	1979	哲学、社会科学、自然科学	91158	17410
化工汉语主题词表	化学工业部科学技术情报所	1983	化工	16602	3075
石油工业汉语主题词表	石油工业部科技情报所	1985	石油工业	6384	515 化学名词 689
国防科学技术叙词表	国防科工委科技情报所	1985	科学技术	29774	4742
铁路汉语主题词表	铁道部科学技术情报所	1987	铁道等	4447	82
新闻叙词表	新华通讯社新闻资料检索研究室	1988	社会科学	8603	1203
艺术科学叙词表	中国艺术研究院	1992	艺术科学	8461	975
中国档案主题词表	该表编委会	1988	综合性	22759	4529
军用主题词表	军事科学院等	1990	军事理论、军事技术	47340	5160

资料来源:[1]侯汉清、马张华:《主题法导论》,北大出版社,1991

[2]有关词表

5.2 二部最具代表性的大型词表

1.《汉语主题词表》

简称《汉表》。中国科学技术情报研究所和北京图书馆主持编制，直接参加编表人 1378 人，历时 5 年，于 1979 年由科学技术文献出版社出版。全表共分 3 卷 10 册，依次为：第 1 卷　社会科学（共二册），第 2 卷　自然科学（共七册），第 3 卷　附表。共收词 108568 条，其中正式叙词 91158 条，非正式叙词 17410 条。1985 年，获国家科学技术进步二等奖。《中国分类主题词表》编制时，对《汉表》陈旧词义进行了删除，增补了 15000 词．因此，《中国分类主题词表》也可看作是《汉表》的增补本。中国科技情报研究所对《汉表》自然科学部分进行了修订，并由科学技术文献出版社出版了《汉语主题词表》（自然科学增补本）。此后侯汉清等人又利用计算机辅助技术为之编制了轮排索引，收录 16 万多条索引栏目，近 600 万字，采用 KWIC 方式，该索引也由科技文献出版社正式出版。

《汉表》是一部大型综合性叙词表，全表由主表、附表和辅助索引三大部分组成。主表就是字顺表，分社科和自科两部分，按汉字的拼音逐字母顺序排列，构成字顺体系。有四个附表，即“世界各国政区名称”、“组织机构名称”、“自然地理区划名称”和“人名”。索引有词族索引、范畴索引和英汉对照索引。

《汉表》是我国编制得比较早的大型词表，由于当时的时代背景，学习国外先进技术不够，使词表编制中存在一些问题，但作为一部词表仍有很多可取的方面，如：它确定了我国传统叙词表的基本模式，并成为典范；为我国词表编制培养了人才，积累了经验；成为各专业性汉语叙词表的词汇和结构的基础。

2.《军用主题词表》

简称《军表》。国家社科“七五”规划基金军事学重点课题，国家军用标准“七五”规划重点项目。军事科学院为主办单位，全军各有关单位参加。军事科学出版社 1990 年 12 月出版。1991 年 8 月由三总部颁发全军使用。《军表》共收词 52500 个，其中正式主题词 47340 个，非正式主题词 5160 个，内容涉及军事科学领域的军事理论、军事技术两部类，87 个学科、专业。该表共分 3 卷，503 万字。第 1 卷主表（字顺表），是词表的主体，也是标引与检索文献的主要工具。该卷共收词 46610 个，按汉语拼音音序排列，并显示每一个主题词的全部词间关系。编有汉语拼音索引，汉字笔画索引。第 2 卷范畴表、附表，是标引与检索文献的辅助工具。其中范畴表是将主表中所收主题词按学科、专业划分排列，供使用人员从范畴分类角度查找主题词；附表由专用主题词组成，共收词 5890 个，包括地名表、人名表和军事著述名称表，可与主表配合使用。第 3 卷汉英——英汉对照索引，将主表和附表中的汉语主题词及对应的英译名，按汉英对照和英汉对照方式排列组成，主要供对外交流军事文献时标引英文主题词和标引英文军事文献时使用，亦可做汉英——英汉军事词典使用。《军表》1992 年获首届全军军事科研成果特别奖和军事科学院科研成果特别奖，1993 年获国防科工委军用标准化科技进步一等奖，1995 年获国家科技进步二等奖。

为了更好地推广和使用《军表》，先后完成了一些配套工具，主要包括：(1)《军用主题词表使用手册》，军用主题词表编委会办公室组织编写，军事科学出版社 1990 年出版，1994 年修订再版。(2)《军用主题词释义词典》，1990 年国家军用标准项目。军用主题词表编管会组织协调有关单位编纂，军事科学出版社 1993 年 10 月出版。全书收词 52234 条，共 235 万字。其作用是使《军表》中的每一个主题词概念准确，词义单一，以便标引、检索人员正确使用，同时兼有规范军事用语的作用。(3)《军用主题词表专题录像片》，是推广使用《军表》的教学片，1995 年完成。(4)军事文献主

题词标引系列规则，由《军事文献主题词标引通则》(GJB2418.1)等五个分标准组成。1992年国家军用标准计划项目，1994年12月完成，1995年5月国防科工委批准颁布，1995年12月实施。(5)军用主题词表应用管理系统(EMT)。

为了满足一些专业部门和单位的需要，编制了一批与《军表》相互兼容，互为补充的专业词表。为了确保与《军表》的兼容，使之成系列，制定了《军用主题词表编制规则》，开发研制了《军用专业词表微机编制管理系统》。在此基础上，全军计划编制的军用公文、军用档案、军用机要、军用气象水文、军事测绘与军用地理、军事情报、技侦情报、技侦技术、通信兵、防化兵、防护工程、军用后勤、军用医学、后勤科技、第二炮兵、军用航空、机场工程、国防科技、武器型号、军队政治工作等20部专业词表已完成编制，并在各系统投入使用。

5.3 主题法编制理论与技术研究

主要有如下方面：(1)叙词表的基本原理的研究。确立概念组配的基本原理，并且综合了体系分类法、组配分类法、标题法、单元词法以及关键法等多种检索语言的原理和方法。(2)叙词表的宏观结构和微观结构的优化。在宏观结构上打破了《汉表》的确立的"五表齐全"式的宏观结构，可根据词表应用的需要选择结构，特别是分类法主题法一体化的对照表、分面叙词表的采用，使汉语叙词表的结构有了新的突破。《国防科学技术叙词表》和《军用主题词表》的等级关系全显示的应用，极大地优化了词表的微观结构。(3)计算机辅助技术的应用，提高了编表效率，保证了编表质量，而且使词表的编、管、用一体化，在九十年代编制的词表都运用了计算机编表技术。(4)词表编制和使用的标准化，制定了一系列标准，使词表纳入了规范化的范畴，词表的标准化也带来了词表的兼容化的进展。(5)叙词表与自然语言的接口及结合使用，以及运用叙词表控制手段编制自然语言检索的后控制词表。

可以预见，叙词语言在一定时期里仍是检索语言的主流之一，其控制技术和方法将得到更加广泛的应用和推广。

6 一体化词表篇

分类法主题法一体化是当今情报检索语言主要发展趋势之一。从60年代初步探索开始，到90年代，我国共发表了100多篇研究和评论文章，编制了近20部一体化词表。一体化词表成为我国情报检索语言的一个重要组成部分，编制技术和理论也走向成熟，在《主题法导论》和《情报语言学基础》等教材中，一体化词表都作为一章写入教材体系中。

6.1 一体化词表的早期理论探索

早在60年代，刘国钧的《分类、标题和目录》和《分类法与标题法在检索工作中的作用——在检索方法座谈会上的发言》两篇文章就曾对分类法和标题法进行了比较。刘国钧认为"分类法和标题法各有所长，各有所短，合其双美，离则两伤。近来分类法和主题法各向对方转变，是很值得注意的一种趋势"。杜定友在《图书分类法的路向》一文中认为："由于分类目录与主题目录都存在着问题，就出现了两种新的方法：活动分类系统和分类主题目录"。杜定友的这一思想在《图书分类主题目录的建议》一文中又得到进一步发挥。而《图书资料分类法》的编制可以看作是他上述思想的实践组成部分。这部分类法由"类质表"和"类素表"组成，"类质表"反映学科分类，"类素表"是反映各种辅助区分、

1976 年丘峰发表《主题法与分类法》一文将分类法和主题法各种类型都进行了比较。1983 年侯汉清发表的《也论分类法和主题法的异同》则注重同时从相同和不同两个方面对分类法和主题法进行比较。对分类法和主题法更系统更深入的比较研究是张琪玉发表的一系列论文和著作。在其《情报检索语言》一书中在论述到每种语言时都设有专节进行比较研究。

1981 年侯汉清在北京大学图书馆学系以题为《论分类法和主题法的一体化》完成了他的硕士论文。在这一几万字的论文中，作者比较系统地探讨了分类法主题法一体化这一课题，并比较全面地概述了国外分类法主题法一体化的研究成果和动态。此后，作者在这一研究的基础上又先后撰文评介了国外几部具有代表性的分类主题一体化词表。如 1969 年由琼·艾奇逊等人在《英国电气分面分类法》（第三版）基础上修订编制而成的《分面叙词表：工程及相关学科的叙词表和分面分类法》，1977 年出版的《联合国教科文组织叙词表》，1981 年由英国标准机构出版的《基础叙词表》等等。《分面叙词表》的功绩在于首次建立了一个真正实现了分类法和主题法两者合一的系统，拟定了方法，建造了模型。而《基础叙词表》由英国标准机构来出版表明分类法主题法一体化的叙词表至少在英国已被作为一种楷模被接受。这些对国外分类法主题法一体化成就的介绍，不仅使我们在一定程度上对分类法主题法一体化的理论和方法有所了解，同时也有助于我们对分类主题一体化必要性和可行性的认识。

6.2 一体化词表的实践探索

一体化词表有多种编制模式，如改造分类表，改造叙词表，编制对照索引，编制分面叙词表等。我国一体化词表编制模式主要是后两种。

1984 年，北京师范大学图书馆学系杨丹曾将《中图法》和《汉语主题词表》中的文化教育部分对照编成了《文化教育主题分类表》和《文化教育分类主题表》。1986 年武汉大学图书情报学院的邓顺国在张琪玉教授指导下完成了题为《分类法主题法一体化问题的探讨以及〈中国图书馆图书分类法〉类目与＜汉语主题词表＞主题词对应试验》的硕士学位论文。作者不仅对分类主题一体化进行了探讨，而且着重在于对《中图法》和《汉语主题词表》对应编表进行了系统研究，分析了编制对照索引的可行性和困难性。邓顺国的这一研究为编制《中国分类主题词表》提供了理论和实践依据

1987 年由北京图书馆图书馆学研究部和武汉大学图书情报研究所共同提出编制《中图法》和《汉语主题词表》相对应的《中国分类主题词表》的计划。与此同时，北京图书馆图书馆学研究部以《中图法》“U6 水路运输”与《汉语主题词表》相应部分进行了对照试验。并且为了实施其编制计划，先后在青岛和重庆召开了编制研讨会。《中国分类主题词表》经过五年的编制于 1993 年完成，之前还有多部专业性对照表编成。如表 6－1 所示：

表 6－1　汉语分类法—叙词表双向对照索引一览表

编号	编制年代	词表名称	主　编	备　注
1	1984	文化教育主题词分类表和分类主题词表	杨　丹	以《中图法》文化教育类为基础编成，由赵燕群教授指导
2	1986	《中图法》艺术及生物科学大类与《汉表》双向对照索引	邓顺国	由张琪玉教授指导
3	1987	《中图法》U6 水路运输与《汉表》双向对照索引	北京图书馆图书馆学研究部	
4	1987－1993	中国分类主题词表	刘湘生、张琪玉等	1994 年华艺出版社出版，共 6 册

编号	编制年代	词表名称	主　编	备　注
5	1990	《中图法》与 MeSH、《中医药学主题词表》对照表	林美兰等	1992 年中国科学技术出版社出版
6	1991	航空航天医学分类主题词表	航空航天医学研究所	
7	1992	《中图法》教育专业分类表与《教育主题词表》对照索引	赵燕群、柳芳等	1993 年由教育科学出版社出版
8		邮电主题词表		范畴索引为《中图法》与词表双向对照索引

资料来源：侯汉清、陈树年：汉语分类主题一体化词表的进展和技术特色，见《索引技术和索引标准》（北京图书馆出版社，1997）。

《中国分类主题词表》从兼顾《中图法》和《汉语主题词表》使用基础来说，是我国分类法主题法一体化非常现实的道路。但是从理论上说，它并不是分类法主题法一体化理想的模式，它受《中图法》和《汉语主题词表》的限制，有不少局限性。从国外编表经验看，分类法主题法一体化最理想的模式是分面叙词表。

1983 年太原机械学院熊兴成以《分面叙词表》为楷模，把《中图法》中的 TJ 武器工业类与《汉语主题词表》中武器军事技术装备方面的叙词，编成《常规武器工业分面叙词表》。这是我国图书情报界编制分面叙词表的首次尝试。

1986 年底在北京召开的全国社会科学情报工作会议决定由中国社会科学院文献情报中心负责筹备组织全国力量编制一部社会科学情报检索语言。会后通过对部分专家的咨询，建议编成一部社会科学分面叙词表。为此，1987 年 3 月在北京黑龙潭疗养院约请全国部分专家对这一方案进行论证，结果也是倾向于编成分面叙词表。会后由侯汉清和戴维民编制《教育分面叙词表》作为编制分面叙词表的示范表。虽然《社会科学检索词表》历时十多年才编成出版，但期间对分面叙词表的编制技术却进行了深入的研究，同时有多部专业性的分面叙词表编成出版并投入应用。如表 6－2 所示：

表 6－2　汉语分面叙词表一览表

编号	编制年代	词　表　名	主　编	备　注
1	1983	常规武器工业分面叙词表	熊兴成	以《中图法》TJ 武器工业类为基础编成
2	1987	教育分面叙词表	侯汉清、戴维民	用吴政机编软件编成
3	1987	航空航天技术主题词分面表	钟　慧	以中图法 V 航空航天大类为基础编成，经张琪玉教授指导
4	1987-1992	农业科学叙词表	方陆明等	1994 年中国农业出版社出版，共 4 册
5	1987-1993	社会科学检索词表	朱铁声、侯汉清等	1996 年社会科学文献出版社出版
6	1987-1988	标准文献主题词表	中国标准化综合研究所	ISO 的 Root Thesaurus 的中译本，共 3 册
7	1987-1990	教育主题词表	曹青阳、赵燕群等	1993 年教育科学出版社出版
8	1990-1991	中国人口主题词表	孟健民、王荣授	1993 年贺玲勇等已制成联机多功能一体化词表

编号	编制年代	词表名	主编	备注
9	1992-1994	音像资料叙词表	赵锦、戴维民	1996年计量出版社出版
10	1992	中国语文教育分面叙词表	胡庆云	
11	1992-1994	服装叙词表	洪漪等	

资料来源:同表6-1

6.3 三部大型汉语一体化词表

两种典型的一体化词表的编制成功,不仅对词表编制模式有创新发展,同时在计算机辅助编表方面也有重大进步。下面三部大型一体化词表都成功地运用了机编技术,极大地提高了词表编制质量,实现了词表的编、管、用一体化。

1.《中国分类主题词表》

《中国分类主题词表》是国家社会科学基金支持的国家"七五"哲学、社会科学研究重点项目。历时八年,于1993年编制完成,1994年北京华艺出版社正式出版。全表约1500万字,共2卷6册。

词表主体由二部分组成,即分类号—主题词对应表(第1卷共2册)和主题词—分类号对应表(第2卷共4册)。第1卷把《汉语主题词表》全部主题词纳入《中图法》的分类体系,既可作为分类法使用,又可通过学科分类途径选类查词。第2卷在增补修订《汉语主题词表》的基础上,构成了18万余条主题词款目,并对应《中图法》号码,是一部主题词表式的《中图法》类目相关索引。因此,《中国分类主题词表》既是《中图法》第三版和《资料法》的合璧本,也是一部完整的《汉语主题词表》的增订本,同时又是一部《中图法》和《汉语主题词表》的双向对应互换、互为索引的工具书,其功能大大强于两者之和。它可以通过标引数据的转换及时完成分类标引和主题标引,实现分类检索和主题检索的互补。

《中国分类主题词表》成为同类型专业性词表编制的基础,是检索语言的创新改革,一体化词表编制模式的成功实践,具有重要的理论价值和实践意义。

2.《农业科学叙词表》

中国农业科学院科技文献信息中心于1990年组织完成了农业叙词库管理系统和编表系统的研制,并建成《农业叙词库》,在此基础上于1992年研制完成了《农业科学叙词表》,1994年农业出版社出版,共4册。《农业科学叙词表》是一部大型一体化词表,共收叙词6.3万条,其中正式叙词5.1万条,非正式叙词1.2万条,并为其配有6万余条的生物名称。分类体系分四大部类,42个一级类目,212个二级类目,174个三级类目,包含60个学科,200多个专业。

词表由字顺结构与分类结构两部分组成。字顺结构包括字顺表、字顺索引、逆序轮排索引、英(拉)汉对照索引和生物名称字顺表五部分,分类结构由叙词分类表和生物名称分类表两部分组成。

《农业科学叙词表》在体系结构上、编制方法技术上都有创新,它以叙词的学科类目整体为单元来编制代、参、分三项关系,字顺表中采用等级关系全显示方式,使其词间关系更加合理。是一部多语种词表,充分实现了与国外著名农业词表的兼容,以及分类语言与主题语言的兼容。编制工作实现自动化,与词库建设、数据库建库一体化。因此,《农业科学叙词表》是一部大型的,结构上有特色,技术上有创新的一体化词表。

3.《社会科学检索词表》

国家哲学社会科学"七五"重点课题，由中国社会科学院文献信息中心主持研制，朱铁声、吕其苏、侯汉清主编。1993年7月20日至1993年8月10日通过书面鉴定，1996年5月由社会科学文献出版社出版。全表共收词23000个，其中正式叙词18874个，非正式叙词4126个。全表由字顺表、分类表和英汉译名对照表三部分组成，是一部大型的社会科学领域的分类法主题法一体化模式的叙词表。

词表在编制初期，经过研究论证和编制试验，选择了分类法主题法一体化模式，词表的分类表和字顺表是两个对应互换的组成部分。分类表分马克思列宁主义、哲学、宗教学、语言学、文学历史学、考古学、社会学、民族学、经济学、政治学、法学、军事学、传播学与新闻学、图书馆学－情报学－档案学、教育学、管理学17个基本大类，26个一级学科类目，220个二级学科类目。根据社会科学学科特点将学科类目划分为历史、学派、方法论、学科、理论、结构、管理、时间和空间面。因此，分类表部分实际上是一部完整的半分面分类叙词表。字顺表是全部叙词按汉语拼音排序，每个叙词与分类表中的类目一一对应，并通过分类号进行连接。

词表不仅填补了我国社会科学领域大型综合性检索语言的空白，而且在探索词表新模式，编制新技术方面进行了大量的理论研究和实践探索，丰富了我国情报语言学理论，我国90年代编制的一批分类法主题法一体化词表都直接和间接运用了《社会科学检索词表》的研究成果。

7 标准化篇

标准化是"在经济、技术、科学及管理等社会实践中，对重复性事物和概念通过制订、发布和实施标准，达到统一，以获得最佳秩序和社会效益。"(GB3935.1－83《标准化基本术语第一部分》)。情报文献标准化是标准化工作的一个重要组成部分。1979年全国情报文献工作标准化技术委员会成立，其中第五分会是主题标引分技术委员会，它的工作任务是：研究并制订文献标引语言有关国家标准，包括研究分类法及其标引规则，制订单语种和多语种词表规则以及文献索引技术等。20年来，情报检索语言标准化及标引标准化取得了一些显著进展。

7.1 标准化进程与成就

建国后，我国分类法的使用出现了不统一局面，1950年6月，文化部郑振铎副部长主持召开的图书分类法座谈会，其目的就是想编制一部能够统一使用的分类法。可以说在以后的几十年中都主要致力于这方面。50年代的《人大法》和《科图法》由于其编制技术的先进，使用的面比较广泛，影响比较大，但还达不到统一分类法的目标。1975年，《中图法》的问世，开创了分类法的新时代，它被各类图书馆广泛使用，从此不再讨论编制新的综合性分类法问题。1980年，《中图法》第二版修订出版以后，使用面更加广泛，有统计说，全国各级各类图书馆有96％使用《中图法》。《中图法》事实上已成为一部标准分类法。1974年，国家为了促进电子计算机在国民经济、情报检索、新闻报导和出版印刷等领域的应用，提出了"汉字信息处理系统工程"(简称748工程)。与之相配套，动员全国力量编制了《汉语主题词表》，该表1980年出版以后，由于篇幅巨大，影响了其推广使用，但其所确立的词表结构体系和词汇基础都得到广泛的认可，为其他词表的编制起到了规范作用，具有标准化意义。

1980年12月全国情报文献工作标准化技术类员会第五分会等在南宁召开了"全国分类法、主题法检索体系标准化会议"。会议的目的是为使全国图书馆、情报和档案系统的文献检索体系逐步趋向统一。会议议题主要有三项：讨论文献检索体系标准化的意义、方针和原则；选定具体的分类

法、主题词表作为国家标准(草案);讨论制定分类法和主题词表(国家标准)的实施方案。会议最后建议选用《中图法》作为国家试行标准草案和在全国试用《汉语主题词表》。根据这一建议,国家标准局于1981年10月下发了关于选用《中图法》作为国家试行标准草案和在全国试用《汉语主题词表》的建议的通知。

从80年代开始,标准化研究和标准的制定进程加快,提出了系列标准草案和送审稿。根据这一背景,文标会五分会等于1989年10月底在成都召开了"全国分类法与主题法标准化研讨会"。会议的目的是推动我国图书情报和档案系统检索语言体系逐步走向标准化。会议的主要内容是讨论有关分类法、主题法的四项国家标准(送审稿),即《文献分类颜色标识规则》、《中图法》作为国家标准送审稿、《文献分类规则》国家标准(送审稿)、《文献叙词标引规则》国家标准(修订送审稿)。

到目前为止,情报检索语言及标引方面的标准,有二项未批的标准报批稿,即:《文献分类标引规则》(报批稿)和《文献分类颜色标识规则》(报批稿),已作为国家标准批准颁布的有五项:

GB/T13745　《学科分类与代码》
GB3860　《文献叙词标引规则》
GB13190　《汉语叙词表编制规则》
GB15417　《文献多语种叙词表编制规则》
GB15418　《档案文献标引规则》

90年代开始,各行业和军队也相应制订了一些标准,其中最具影响的是关于军用主题词表编制规则和军事文献标引规则的国家军用标准,具体包括:

GJB1776—93　《军用主题词表编制规则》
GJB2418.1　《军事文献主题词标引规则》
GJB2418.2　《军队档案主题词标引规则》
GJB2418.3　《军队情报主题词标引规则》
GJB2418.4　《军事图书主题词标引规则》
GJB2418.5　《军队机关公文主题词标引规则》

在标准化方面,长期研究和呼吁的是关于书次号的标准化,许多学者进行了积根倡导。实现书次号编制标准化最有效的途径就是编制统一使用的著者号码表。但我国建国后,各自为政,编制了几十种著者号码表,其中主要的如表7-1所示:

表7-1　主要中文著者号码表

序号	著者号码表	编者	出版者	出版年	页码	开本
1	万氏著者号码表	万国鼎	北京图书馆	1956	24	16
2	笔画字顺著者号码表	王风翥	北大图书馆学系	1954	27	16
3	汉语拼音著者号码表	刘德人	北京矿业学院图书馆	1958	88	16
4	汉语拼音著者号码表	袁涌进、周树基		1959	90	16
5	汉语拼音著者号码表	武汉大学图书馆	编者	1974	64	16
6	中文图书汉语拼音著者号码表	李修宇等	哈尔滨船舶工程学院出版社	1979	18	16
7	蒋完奎著者号码表	蒋完奎				
8	通用汉语著者号码表	刘湘生主编	海洋出版社	1992	110	16

《通用汉语著者号码表》是由文标会五分会组织编制的,主要是为了促进我国中文图书同类书

排列书次号的规范化、标准化。是一部编制得非常实用的著者号码表。但各图书馆在书次号编制上都有自己的习惯做法，因此，《通用汉语著者号码表》推广也是有一定难度的。

外文图书采用著者号码表基本趋向统一，英文图书采用《克特著者号码表》，俄文图书采用《俄文图书著者号码表》，日文图书采用《植村著者号码表》等。

在我国标准化进程中，也做了大量引进和介绍国际标准和其他国家标准工作，以对我国标准制订起指导和参考作用。

UDC作为全世界广泛使用的分类法，实际上是一部国际标准分类法。有些国家规定在一些出版物上需标注UDC号码。我国GB7713－1987规定，科学技术报告、学位论文、学术论文的封面上应尽可能注明UDC号。ISO规定，凡该组织及其成员国颁布的标准文献上必须印UDC类号，我国也已执行该决议。《国际专利分类法》是专利领域的一部最通用的分类法，是实际上的标准分类法。英国标准学会为ISO编制的《基础叙词表》，企图以标准形式确立其标准性质，我国也翻译了这部一体化词表，名称为《标准文献主题词表》，专门用于标准文献的标引。我国引入的国际标准主要有：

ISO2788—1986 文献工作——单语种叙词表编制与修订准则

ISO5693—1985 文献工作——文献审读、主题分析与选定标引词的方法

ISO5694—1985 文献工作——多语种叙词表的编制与修订准则

7.2 标准化与兼容化

标准化和兼容化是密不可分的，因此，研究标准化必然要研究兼容化的问题。

印度尼莱姆汉(Neelamenham,A.)曾指出："为了使从不同角度建立的情报系统便于统一与协调运行，如果各个系统表达主题所用的'语言'在语法上是一致的，彼此是兼容的，而且以合理的费用可以互相转换，那么将是有益的。"

所谓兼容，是指两个实体结合起来工作的能力。具体地说，情报检索的兼容性是指某种检索语言及其构造的检索式(或标引记录)，可以直接适应于或通过交换适应于多个情报检索系统。

关于检索语言兼容问题的研究，国外开始于60年代。1962—1965年期间，美国的一些学者曾对美国一个政府机构所用的检索语言转换为另一机构的检索语言的难易程度进行了多次研究，试图在几个机构的词表中建立一个等价词的主表，以便用于编制联合主题目录。1971年10月，世界科学技术情报系统(后改名为综合情报计划)召开的由111个国家、62个国际组织参加的大会十分重视这一问题。会议通过的报告指出："需要研制一些更好地控制和转换自然语言和情报检索语言的工具"。格卢斯科夫(Gluskov)等学者对检索语言的兼容问题进行了深入研究，还分出结构兼容性和语义兼容性。结构兼容性又可归纳为形态兼容性和句法兼容性。语义兼容性可归纳为词汇的、词形的和语段的兼容性。

美国情报学家兰开斯特曾为联合国教科文组织撰写了题为《影响情报系统和情报机构的兼容性问题》的报告，这份报告后来被收入《情报系统的兼容性》和《情报检索词汇控制》(第二版)两书中，该报告全面概述了国外关于兼容性问题研究的全貌和主要成就，分别介绍了词表的兼容性、中介词典、集成词表方法、微观词表、宏观词表等有关问题。上述两书均被译成中文介绍到我国。

国外在兼容性方面也有许多成功实践，世界科学技术情报系统计划项目《概略分类体系》(BSO)就是宏观词表的典型示例，BSO也被翻译成中文，并一度成为研究和关注的热点。

曾蕾在80年代初就系统地研究了检索语言兼容化的问题，直到90年代共发表了《检索语言兼容化的趋势》、《建立我国统一的叙词语言体系》和《联机多数据库检索环境中保证检索语言兼容能

力的手段》等一系列论文。1987 年武汉大学洪漪完成了题为《汉语叙词语言的兼容问题研究》的硕士论文，这是国内对汉语叙词语言兼容问题的系统研究。侯汉清 1989 年发表《叙词表的兼容与互换》一文对国外兼容性问题研究成果进行了系统介绍和研究。

最近于 1996 年张琪玉在《情报理论与实践》上发表《情报检索语言的标准化与兼容化》，对标准化与兼容化的关系，以及兼容的理论和方法问题进行了更加系统深入的研究。在这一篇文章中，作者将兼容模式分为主动兼容和被动兼容模式，对每种模式进行系统归纳，如表 7－2 所示：

表 7－2　情报检索语言的兼容模式

主动兼容模式	被动兼容模式
制定词表编制规范模式 系列化分类表或词表模式 微型词表模式 编者选词或列类时参考权威性词表或分类表，并尽可能与之保持一致模式 分类法主题法一体化标引工具模式 通用词表模式 词表编制时基本依据国外几种主要专业词表模式 直接利用国外的词表，仅根据国情作适当补充的模式 词库模式	对应表模式 　两种语言对应表模式 　中介词典模式 集成词表模式 媒介语模式 入口词表模式

资料来源：张琪玉：情报检索语言标准化与兼容化，《情报理论与实践》，1996(5)

与标准化相同步，我国情报检索语言兼容化在实践上也取得了丰硕成果。主要表现在以下几个方面：(1)通过制定标准，来规范词表的结构模式和选词标准，如叙词表编制的国家标准和国家军用标准；(2)《中图法》和《汉语主题词表》等综合性检索语言由于其权威性和影响性，同专业性的分类表和词表在实践上实现兼容；(3)通过编制对应表式一体化词表和分面叙词表式一体化词表，实现两种语言的兼容；(4)在一个系统内，通过有效的组织和提供统一的计算机辅助编表系统，实现系统内部兼容，如《军用主题词表》与 20 部专业词表实现了有效的兼容，《国防科学技术叙词表》在编制机读版时，将 11 部词表集成起来，实现兼容。

标准化和兼容化是两个密不可分的问题，无论是过去，还是现在都有重要的研究意义。在现今网络环境下计算机信息检索，情报检索语言兼容化将有许多新的研究课题，许多实践问题需要研究和解决。

8　技术篇

计算机技术的应用是情报检索语言及标引与检索技术取得重大技术进步的基础。因此，技术篇所述就是特指计算机技术应用于情报检索语言管理和联机环境下自然语言检索及相关技术。

8.1　分类法和词表的计算机辅助编制与管理

国外在 60 年代就已成功地运用计算机技术辅助编制和管理词表。1966 年哈蒙德(Hammond, w.)与 ARIES 公司设计了叙词表编制和管理计算机程序，并应用于《NASA 叙词表》和《水源叙词表》的编制工作。进入 70 年代，尤其是 80 年代，计算机编制和管理词表已是一项成熟技术，并编制出一大批词表。至 80 年代对国外 227 部叙词表的统计，机读版词表有 79 部，占 34.8%，超过了三分之一。而到 90 年代，机编分类表和词表已相当普及，一些传统的分类表和词表也运用计算机技

术进行了改造,如《杜威法》就经历了电子版、视窗版和网络版三个阶段。词表的联机显示和网络版使受控语言达到了新的应用阶段。

我国70年代开始探讨叙词表的计算机编制和管理技术,但直到80年代才有应用成果。中国科学院计算技术研究所从1983年10月开始研制词表管理系统(TMS系统),用于编制和管理《计算机科学技术英语词表》。1987年在南京和上海成功地用计算机软件在IBM/PC微机上编制了两部汉语叙词表——《教育分面叙词表》和《教育科学叙词表》。后来的《教育主题词表》、《中国分类主题词表》、《社会科学检索词表》、《中国人口主题词表》、《音像资料叙词表》、《军用主题词表》及其系列专业词表、《农业科学叙词表》、《国防科学技术叙词表》等都是运用计算机技术编制成功的。有些在编表的同时建立了比较完善的词表管理系统,其中最具代表性的有《军用主题词表》、《农业科学叙词表》和《国防科学技术叙词表》管理系统等。

由文化部立项,山西省图书馆与北京图书馆、中国科学院文献情报中心、中国人民大学图书馆研制的《计算机文献标引对照系统》于1997年完成并通过鉴定。它在《中国分类主题词表》基础上予以扩展,实现了《中图法》、《科图法》和《人大法》号码的自由转换,以及三法的号码与《汉语主题词表》的主题词相对应,具有联机标引功能,也可用于检索选词选号。

下面择要介绍的三个系统各具一定的代表性。

1.《军用主题词表》应用管理系统(EMT)

在《军表》编制时就成功运用了计算机辅助技术,专业表编制时开发了军用专业主题词表微机编制管理系统,实现了编表、词表质量控制、排版、动态管理一体化。

在此基础上,全军军表编管会办公室和总参三部九局主持研制了EMT,1995年初至1998年初共历时三年研制完成。1998年9月25日通过技术鉴定。

EMT具有以下技术特点:(1)具有一般机读词表的显示浏览、自动查询基本功能,还具有辅助标引以及对词表使用情况自动记录和统计分析功能,实现了《军表》应用与自身动态管理一体化;该系统除《军表》外,还收入了《军用主题词释义词典》、《军用主题词表使用手册》、《军事文献主题词系列标引规则》等配套工具,实现了《军表》与其配套工具应用的一体化。(2)采用关系数据库理论和超文本技术,创造了叙词表语义关系联动显示浏览新方式。即对词表各索引表中的主题词均可实现多表语义联动显示,一次为用户提供所查主题词全方位语义信息。(3)支持用户在标引中使用自然语言,在实现自然语言与词表中受控主题词自动转换中,采用了智能化的词素相似度模式识别技术,即通过系统构造的含有8万多条词素,近2万对词素对应关系,20多万个转换人口和若干转换规则的知识库,分析自然语言与词表中主题词所含词素的相似程度,自动推荐与自然语言语义相同或相近的主题词,辅助用户完成标引。(4)可对用户标引和检索中使用的主题概念、主题词、自由词及建立的主题概念与主题词对应转换关系进行自动记录和分类统计分析,以量化对比方式跟踪分析用户使用词表的情况,并能将根据统计分析获取的相关数据,具有一定"傻瓜"性能,终端用户只要了解文献标引的简要知识便可使用,大大降低了对用户的专业技术要求;响应速度快,各项功能实现均在秒级范围内;使用方式灵活,既可独立使用,也可插入检索系统或在网络上使用。

2.中国农业叙词库及其管理系统

中国农业科学院科技文献信息中心与中国人民大学信息中心等单位共同研制。系统由6个子系统组成:(1)叙词数据录入、转换、建库子系统;(2)计算机辅助编表系统;(3)叙词库输出系统;(4)叙词库维护系统;(5)叙词库用户系统;(6)叙词库开发系统。系统建立了中国农业叙词库,由叙词库、属分关系库、用代关系库、参照关系库、汉英关系库、汉拉关系库、汉字属性库7个库组成。系统实现了多项功能,包括:词表的辅助编制功能,对叙词三级分类系统、叙词词间关系建立维护的功

能，对系统的维护功能，标引人员和检索人员的查询功能，编排输出各种格式叙词表的功能。

中国农业叙词库及其管理系统是一个有效且实用的系统，是一个成功的范例。

3.《电子版国防科学技术叙词表》

中国国防科技信息中心组织研制。在国防科技情报系统内使用的11部叙词表的基础上，进行兼容性处理，并增补新词，于1997年编制完成，1998年又进行了修订。收词80005条，其中无关联正式词1672条，非叙词22405条，入口率为43%，关联比为97%。现该表装人了中国国防科技信息中心的计算机辅助主题标引和建库系统，为实现国防科技情报系统各单位信息的协同处理、数据库联网和终端用户上网检索奠定了坚实的基础。电子版词表功能齐全，超越了普通叙词表机读版的功能，而且增挂了叙词导航、定位和叙词范畴选择三个模块来更好地实现辅助标引功能，实现了叙词表编、管、用一体化的模式；词间关系准确、合理、完备，词族等级深，允许双向成族，一个叙词最多归入了6个上位类，参照词最多达到60个；显示方式灵活，既可按全显示方式进行显示，也可按用户需要分别显示部分词间关系。

8.2 自动标引

自动标引是美国IBM公司卢恩等人开创的一个研究领域。1956、1958年卢恩先后撰文提出自动抽词标引的基本思想：一篇文章中一个词出现的频率是这个词的重要性的有效测度。一个句子中具有给定重要测度的词的相关状态，成为该句子重要性的有效测度。按照词的出现频率，以一定的标准排除高频词与低频词，剩下的就是最能代表文献内容的词。

卢恩提出的基于词频统计的标引方法一直到60年代前期都是自动标引采用的主要技术。从70年代后期到70年代末，自动标引研究取得了很大的进展，提出了概率统计标引法、句法分析标引法以及各种加权模型等。并且进行了大量的自动标引试验及评价，建立了一大批应用与试验系统。

80和90年代，由于自然语言检索的盛行，关于自动标引研究则更加深入。

我国早在1963年就开始介绍国外自动标引研究情况，但是直到80年代，汉语自动标引研究才开始起步。1980年11月至1981年4月，陈培久利用TK—70计算机和T4100汉字系统建立了"汉语科技文献自动标引试验系统"。在80年代和90年代，国内学者提出许多自动标引方案。由于汉语标引的最大的难题是分词问题，因此许多研究的核心都基于研究汉语分词技术。这些方案从类型上划分有形式匹配法、切分标记法、统计法、语法语义分析法、神经网络法，如表8-1所示：

表8-1 汉语自动标引方案

类 别	技术方案	研制者	成果发表时间
形式匹配法	词典切分组词法	陈培久	1983
	部件词典法	王永成	1984
	OM自动分词法	梁南元	1985
	主题词表法	北大图书馆学系	1987
	关键词表法	邓钦和、毛玉姣	1987，1989
	汉语词素自动分词法	江孝感	1989
	语词结构类比法	赵宗仁	1991
	CWSAIS法	陈 豫、曾民族	1991
切分标记法	前后位切割标志法	姚天顺	1986
	非用词后缀表法	吴尉天	1987
	链接表法	周依欣、吴尉天	1990

类别	技术方案	研制者	成果发表时间
统计法	统计分析分词法	龙泽云	1987
	上下文比较标引法	夏　海	1990
语法语义分析法	词境相关自动分词法	黄祥喜	1989
	生成—测试自动分词法	黄祥喜	1990
神经网络法	神经网络系统法	贺前华、李　粟	1990,1992

资料来源:[1]张琪玉:自然语言检索研究进展,《知识信息管理研究进展》,武汉大学出版社,1998

[2]高崇谦:我国科技文献计算机标引研究的现状,《情报学报》,1989,8(1)

8.3 自动分类

自动分类研究开始于50年代末,现已达到实用化阶段。自动分类包括自动聚类、自动归类和类号的自动转换三部分。1982年侯汉清与黄刚发表《电子计算机与文献分类》探讨和介绍了国外自动分类的研究情况。到80年代中期及其以后,我国有不少学者对自动分类进行了试验和系统开发。如:朱兰娟在王永成指导进行的试验研究,1995年金巍在张琪玉和王永成指导下完成的以肿瘤学专业文献为例的中文自动分类系统研究,1995年苏新宁、徐进鸿、史九林合作研究的档案自动分类方法,1995年叶新明关于基于《中图法》的中文文献自动分类研究等。表8-2是我国研制和试验的主要自动分类系统。

表8-2　我国研制的自动分类系统

类型	研制人	单位	完成时间	主要技术特点
辅助归类系统	莫少强	广州中山图书馆	1984	人工主题分析、系统完成查表、自动归类及附表调用
自动归类系统	朱兰娟等	上海交通大学	1986	类主题词表,检索类主题词累计其类归属度,据 Bayes 最小损失原则确定分类
辅助归类系统	张炳恒	天津医学情报所	1989	将分类法类名分解为单元词,人工主题分析及单元词组配,系统确定类号
自动归类系统	苏新宁等	南京大学	1995	主题词与类号关系表;确定权重系数;分类前控词典;停用词表
自动归类系统	叶新明	杭州应用工程技术学院	1995	类名主题词表;组配词表;非用词表;匹配算法;二字先行,单字进或退
自动归类系统	吴　军	清华大学电子工程系	1995	以语料相关系统作为分类依据;字频、词频及其常用搭配为补充;停用词表;人工指导分类
自动归类系统	王永成 张　坤	上海交通大学	1997	部件词典技术;自动分类用关键词分类归属表
自动归类系统	刘开瑛等	山西大学计算机系	1997	三维加权算法,分词采用最长匹配算法,类别词加权,语料中抽词

资料来源:成颖、史九林:自动分类研究现状与展望,《情报学报》,1999,18(1)

8.4 自然语言检索技术研究

自然语言检索是情报检索发展的一个重要趋势,我国自然语言检索技术研究已出现了一些较为实用的检索系统,理论研究也迈上了一个新台阶,1997年张琪玉教授的《情报语言学基础(增订二版)》将自然语言检索作为一个重要章节写入该书,系统总结了我国自然语言检索技术研究成果,标志着我国的自然语言检索已进入了一个新的历史时期。因特网已席卷全球,自然语言查寻更是

网上信息资源检索的主要手段，检索环境、检索要求的变化，也使自然语言检索技术研究成为当今研究的一个热点。

自然语言检索本身有许多优点，在时间紧，而又缺少标引力量和充足投资的情况下，快速开发数据库新产品，可首先考虑自由标引，建造自然语言检索系统。1997 年 7 月由解放军报社与深圳信源电子有限公司等多家单位共同研制完成的《解放军报光盘数据库系统》是我国存储量最大，内容最完整、最集中的军事信息全文数据库。是我国数据库自由标引的典范，其分类主题标引在报纸标引中具有开拓性意义，数据加工处理、检索的某些性能指标已达到国内和国际先进水平。

全文检索是自然语言检索的重要方式，单汉字检索是汉语全文检索系统中数据处理的主要方式。汉语全文检索系统由单机走向网络，功能更为完善，我国全文检索系统中以北大方正和东方龙马开发的系统最为出名。世界流行的全文检索模型有布尔逻辑检索模式、概率推理模型和空间向量模型。我国的检索模型单一，主要以布尔逻辑检索为主，各类截词检索是最为常见的方式，概率推理方式和空间向量方式，理论研究少，实用的检索系统更少。

后控制词表理论研究日臻成熟，对我国词表的功能、形成等问题的研究带来了新的视角。自然语言检索做为情报检索的一种方式，离不开情报检索的基本原理，后控制词表作为对自然语言的有效控制手段，为改进联机网络内的检索效果起着重要作用。专家学者一致认为后控制词表是实现受控语言与自然语言结合的有效途径，并对其控制机理、控制程序、检索效率和词表编制方法展开了广泛的讨论。后控制词表的编制一般以原有的自然语言检索标识为基础，通用性和易用性不高，侯汉清曾主张以《中国分类主题词表》为基础，建立后控制词表，应用于自然语言检索系统，以实现分类主题一体化词表与自然语言的结合，增加检索语言与自然语言的对应转换。

自然语言接口是叙词表中自然语言检索的入口，性能良好的自然语言接口能较好地实现自然语言与受控语言的转换，提高自然语言与叙词语言的兼容度，提供多种查检途径。1998 年通过鉴定的“《军用主题词表》应用管理系统”提供的自然语言接口是通过对叙词词素的轮排来实现的，并加入了人工智能的原理和方法，在国内处于领先地位。

搜索引擎成为新一代自然语言检索工具。搜索引擎是适应国际互联网上信息资源的发展应运而行的。国内搜索引擎也发展到了几十种，按检索方式的区别可分为分类目录式搜索引擎和主题索引式搜索引擎。国内分类目录式搜索引擎以“搜狐”为代表，主题索引式搜索引擎以北大的“天网”为代表。目前的搜索引擎检索机理很不完善，大多数只能提供简单的关键词检索，检全率和检准率均不高，如何把后控制词表技术引入到搜索引擎的编制中是一个急需解决的问题。

与国外自然语言检索研究相比，我国还有很大的差距。自然语言检索所面临的两大难题——一是如何从自然语言文本中自动抽出最能准确、充分地表达文献主题的词，以及这些词与检索课题有效匹配的问题；二是克服自然语言由于不规范和缺乏语义关联性而对检索不利的问题。

9 未来篇

21 世纪已向我们走来，网络化使信息流更加泛滥，信息检索和利用的困难和矛盾将更加突出，作为情报检索语言保证的检索将朝着什么方向发展，情报语言学将面临什么样的研究课题都是我们必须要思考和深入探索的问题。

9.1 21 世纪情报检索语言的发展方向

从 80 年代后期，尤其是 90 年代，许多学者对未来情报检索语言进行了展望和构想。

曾蕾认为在联网环境中情报检索语言有四个发展趋势，即：情报检索语言词汇的联机显示；多数据库查询中的词汇兼容；采用自由词以提高联机状态下的主题途径查找效率；加强检索语言及检索系统的易用性和透明性。她在参与美国凯斯西部保留大学电子图书馆研究项目中提出一种以高度专业化的知识库为基础的检索语言模式。其设计思想是：允许用户在检索提问中使用自然语言；支持标引、查寻、检索、浏览、组织情报全过程。在系统外部，用户既可以用自己的语言（自然语言），也可用自己熟悉的分类体系查找和组织情报信息，用户并不需要看见和直接使用受控语言。然而在系统的内部，存在着的是以知识库为基础的，高度专业化的受控检索语言，对用户的自然语言检索提问的查寻是由内部受控语言支持的。

《图书馆杂志》1996年第4期发表后来收入上海科技文献出版社出版的《21世纪图书馆展望——访谈录》的吴建中与张琪玉、侯汉清的对话，均描述了情报检索语言未来的发展趋势。两人均认为自然语言检索是情报检索领域的重要发展趋势，但目前在情报检索中情报检索语言仍处于主流地位，自然语言还不可能全面取代情报检索语言。但要重视对自然语言的研究。

张琪玉在1997年分别在《图书馆杂志》和《北京大学学报》（信息管理系专刊）中发表文章，披露了他关于未来新型情报检索语的设计思想，即学科——事物概念——组配型检索语言，它的主要原理和方法是：分面分析＋概念代码＋概念对应转换＋数据库技术。具体地说，它采用概念代码形式，将分类号、概念词以及自然语言词等作为概念代码的各种索引，并实现语词词素的全面轮排，利用该索引语言可从概念分类索引、概念词索引、语词轮排索引入手标引文献，可将概念分类索引和语词轮排索引作为检索入口。是一种集各类情报检索语言以及自然语言之所长的新型检索语言模式。

白国应认为21世纪中国检索语言主要有10个发展方向，即多样化（从单一化走向多样化）、一体化（分类法主题法一体化）、兼容化、国际化（通用化）、组配化、规范化（标准化）、自动化、自然化（采用自然语言）、专业化和整体化。

我们认为，在21世纪，情报检索语言仍是发展主流，并将广泛应用。同时自然语言应用范围将越来越广，并在应用中充分吸取情报检索语言的原理和方法，走相互结合、相互融合的发展之路。在情报检索实践中，情报检索语言将不断优化、改进、探索出新型的理想的情报检索语言模式。

9.2 21世纪情报语言学研究的新课题

80年代后期，美国华盛顿大学的 Raya Fidel 提出了一份关于情报的组织、标引与检索的未来研究课题的建议。曾蕾将这一建议编译后介绍到国内。并在另一篇文章中，对中国检索语言的研究与发展方向提出了三点建议：（1）认识到情报的组织、标引、检索、使用与再使用是一个不可分割的过程，任何设计和研究都应考虑全过程，词表编制不能独立于全系统和全过程之外。（2）情报检索过程中终端用户将从各方面直接介入，因此检索语言必须以终端用户，而不仅仅是标引人员作为词表使用对象考虑。（3）在联网环境下所提出的问题不仅仅在于如何提供更好的检索途径，而且在于如何提供以前从未出现过的检索途径和控制手段。

张琪玉在1997年发表的《世纪之交中国情报语言学发展之路》一文中，提出了我国情报语言学面临的八个方面的研究课题：（1）如何使情报检索语言适应新的检索环境的研究；（2）自然语言检索研究；（3）与知识分类和术语学的结合研究；（4）体系分类法和叙词法分面化及自然语言化的研究；（5）标准化、兼容化和一体化的研究；（6）用户对各种语言使用过程的研究；（7）与国际接轨问题的研究；（8）情报检索语言结构新模式的探索研究。

1998年张琪玉在《情报语言学领域亟待研究且潜藏较富的课题》一文中，更加明确地提出了我

国情报语言学未来需要研究的五个方面15个研究课题，它们是：(1)情报检索语言研究课题，包括体系分类法的改造，叙词表的分面化改造及与自然语言相结合，自然语言接口及对应转换词典；(2)自然语言检索研究课题，包括自动抽词及抽词词典，单汉字检索系统的改进，全文检索系统的改进，后控制词表编制的自动化；(3)网上资源检索研究课题，包括适用于组织和检索网上资源的检索语言的探索，中外主要情报检索语言之间对应转换，中外主要自然语言之间的对应转换；(4)情报语言学基础研究课题，包括知识分类的微观立体网络化，情报语言学与术语学的结合，未来情报检索语言的探索，用户研究；(5)文献标引和检索某些具体问题的研究课题。

应该说，上述提到的未来情报语言学和研究课题是非常丰富的。其中有些是传统的研究课题，表明情报语言学继续原来的发展轨迹朝越来越成熟的方向发展。另一方面，是面临新环境、新情况、新技术而不断产生新的研究课题，这表明情报语言学仍然是一门新兴的学科，具有广阔的发展前景。

参考文献：

[1]张琪玉，丘峰. 我国情报检索语言四十年文献简述. 见《情报检索语言论文选》，书目文献出版社，1990
[2]孙勇，王秀玲. 我国情报检索语言著作研究. 中国图书馆学报，1997(5)
[3]白国应. 走向21世纪的中国文献分类学. 中国图书馆学报，1995(3)
[4]张琪玉. 情报语言学文献库(电子版). 武汉大学出版社，1998
[5]北京图书馆图书馆学研究部. 全国中文图书馆学情报学档案学书刊资料联合目录(草稿)，1987
[6]翟凤岐，张芝兰. 我国主题检索语言的理论与实践(1949－1991)概述. 中国图书馆学报，1993(2)
[7]吴秀爽，赵良英. 近10年我国主题法研究论文的调查与分析. 图书情报工作，1993(1)
[8]侯汉清，王荣授主编. 图书馆分类工作手册. 中国科技技术出版社，1992
[9]戴维民. 中国现代图书分类法. 见《图书馆分类工作手册》，中国科技出版社，1992
[10]戴维民.《中国图书资料分类法》的发展历程和路向. 图书情报论坛，1991(4)
[11]白国应. 中国文献分类学研究的方向、重点及热点. 大学图书馆学报，1995(2)
[12]白国应.《中国人民大学图书馆图书分类法》的历史、功绩、问题和展望. 图书情报论坛，1998(3)
[13]白国应. 中国图书分类学四十年. 高校图书情报学刊，1989(1)
[14]白国应.《科图法》的回顾与展望. 图书馆杂志，1990(3－5)
[15]李兴辉. 中小型图书馆图书分类表草案(评介). 图书馆分类工作手册，中国科技出版社，1992
[16]李兴辉. 大型图书馆图书分类法(简介). 见《图书馆分类工作手册》，中国科技出版社，1992
[17]王贤甫主编. 农业信息·情报理论与实践. 中国农业科技出版社，1993
[18]戴维民. 汉语叙词表的发展现状与未来. 大学图书馆学报，1989(4)
[19]侯汉情. 论汉语叙词表编制的新模式和新方法. 图书情报工作，1990(3)
[20]侯汉清，陈树年. 汉语分类主题一体化词表的进展和技术特色. 见《索引技术和索引标准》，北京图书馆出版社，1997
[21]戴维民. 我国分类主题一体化研究进展. 大学图书馆通讯，1987(6)
[22]张琪玉. 情报检索语言标准化与兼容化. 情报理论与实践，1996(5)
[23]侯汉清. 叙词表的兼容与互换. 文献工作研究，1989(1)
[24]傅兰生. 我国词表工作近期发展方向——兼论建立国家主题词库的可能性. 情报理论与实践，1989(2)
[25]傅兰生. 我国叙词兼容两大方案的分析——兼论国家级叙词兼容词库的建立. 情报学报，1991，10(4)
[26](美)兰开斯特. 情报检索系统的兼容性. 科技文献出版社，1989
[27](美)兰开斯特. 情报检索词汇控制. 同济大学出版社，1992
[28]张琪玉. 情报检索语言的发展趋势——与吴建中的对话. 图书馆杂志，1996(4)
[29]侯汉清. 从人工语言到自然语言——与吴建中对话. 图书馆杂志，1996(4)

[30]戴维民.自然语言标引与检索现状与趋势.情报业务研究,1991,8(4)
[31]曾蕾.联机环境中的情报检索语言.书目文献出版社,1996
[32]洪漪.我国信息网络建设中的检索语言问题.中国图书馆学报,1995(3)
[33]洪漪.我国情报检索语言学学科建设及其今后的研究课题.情报业务研究,1993,10(1)
[34]张琪玉.世纪之交中国情报语言学发展之路.1997年理论学术年刊(《图书馆杂志》1997年增刊)
[35]张琪玉.情报语言领域亟待研究且潜藏较富的课题.1998年理论学术年刊(图书馆杂志1998年增刊)
[36]张琪玉.探索21世纪的情报检索语言.北京大学学报(信息管理学专刊),1997
[37]张琪玉.学科—事物概念组配型检索语言.图书馆杂志,1997(2)
[38]白国应.中国检索语言发展方向.1997年理论学术年刊(《图书馆杂志》1997年增刊)
[39]曾蕾.联网环境中情报检索语言的发展趋势与中国未来情报检索语言的研究与发展路向.情报学报,1994,13(6)
[40]侯汉清,徐佳.国外叙词表的概况及发展趋向.情报学报,1989,8(5)
[41]禹崇谦.我国科技文献计算机标引研究的现状.情报学报,1989,8(1)
[42]侯汉清,黄刚.电子计算机与文献分类.计算机与图书馆,1982(1)
[43]成颖,史九林.自动分类研究现状与展望.情报学报,1999,18(1)

九十年代我国情报语言学研究进展

空军政治学院信息管理系　戴维民　傅　亮

[提要]本文系统概述了我国九十年代情报语言学理论和实践研究进展，包括情报语言学学科理论，分类检索语言，主题检索语言，分类法和主题法的标准化、兼容化、一体化、计算机化，自然语言检索技术等，展望了面向二十一世纪情报语言学的发展方向和研究课题。

[关键词]情报语言学　情报检索语言　自然语言检索　研究进展　综述

1　引言

情报语言学是研究情报检索中语言保证问题的一门学科，其主要研究对象是情报检索语言和自然语言在情报检索中的应用。80年代，我国分类检索语言和主题检索语言都有重大发展，特别是叙词语言建设取得了令人瞩目的成就，与情报检索语言实践紧密相联的理论研究也进入了一个比较繁荣的时期，情报语言学学科体系初步形成，其学科地位得到基本确立。在此基础上，90年代的情报语言学理论和实践都进一步走向成熟，取得了一批重要的研究成果。在新世纪到来之际，总结90年代我国情报语言学的成就，对加强面向21世纪的情报语言学研究具有非常重要的积极意义。

2　情报语言学理论研究的繁荣与学科的趋于完善

90年代，我国情报语言学已基本走向成熟，学科地位得到普遍公认。1997年增订出版的张琪玉教授的《情报语言学基础》将自然语言检索问题和分类法主题法一体化检索语言作为两章纳入情报语言学学科体系，标志着我国情报语言学学科体系进一步完善和情报检索语言研究领域进一步拓展。

90年代，我国情报语言学界同仁对本学科各个领域和方向进行了全方位的研究和探索。笔者依据张琪玉教授的《情报语言学文献库》(电子版)中的目录数据按年份进行了统计。我国情报语言学研究保持了80年代形成的繁荣局面，平均每年都有5、6百篇文献发表或出版。可以预计，到今年年底，90年代出版、发表的文献总量将超出80年代。

情报语言学著作作为情报语言实践理论化、系统化的产物，是衡量一国情报语言水平的重要标志。据统计，90年代出版的著作有40部以上。其中有一批相当具有分量，综合性的情报语言学重要著作除张琪玉的《情报语言学基础(修订二版)》(1997)之外，还有侯汉清的《当代分类法主题法索引法研究》(1993)，戴维民的《情报检索语言综论》(1994)，张琪玉等编的《情报检索语言论文选》(1990)，侯汉清主编的《索引技术和索引标准》(1997)，曾蕾编著的《联机环境中的情报检索语言》(1996)，等等。

作为情报语言学实践成果的各类词表(包括分类表、叙词表、一体化词表等)在理论研究和实际编制、修订技术上都有了重大的突破和进展,三大图书分类法的修订出版标志着我国传统分类法的日臻完善,《中国分类主题词表》、《社会科学检索词表》的编制与出版标志着我国分类法主题法一体化词表技术达到应用水平,《军用主题词表》及其系列化专业词表等数十部叙词表的出版,标志着我国叙词语言体系已基本建立。

90 年代,我国情报语言学研究涉及面宽,重点突出。无论是传统的体系分类法,还是作为当今主流语言的叙词法,是规范化语言还是自然语言,是先组式语言还是后组式语言,是手工检索还是计算机检索,都是研究的对象。当然,研究是有重点有倾向的,主要体现在:分面组配原理和方法的研究;分类法主题法一体化词表的研究;情报检索语言兼容化和标准化研究;规范语言和自然语言结合的研究;情报检索语言计算机化、网络化的研究;自然语言检索法的研究;传统情报检索语言的改造和新型情报检索语言创制的研究;情报检索语言使用问题、易用性问题以及情报检索语言知识普及教育问题的研究;情报语言学基础理论的研究等方面。

90 年代我国情报语言学研究成果展示了学科发展的如下趋势:大型综合性分类语言和系列化的叙词语言是当代情报检索语言的主流;分面分析原理在情报检索语言中得到广泛运用;大型综合性情报检索语言和专业化、专题化语言受到同样的重视,并逐步走向兼容化、标准化、系列化和配套化的发展道路;各种类型情报检索语言朝着相互结合和相互渗透的一体化方向发展,尤其是分类法主题法一体化,受控语言与自然语言相结合使用等。在以上这一系列的趋势当中,计算机技术的运用起到了重要的技术推动作用,因面情报语言的计算机化是当代情报语言学发展中的又一个重要趋势。

3 分类检索语言理论与实践进展

3.1 关于分类检索语言理论和发展方向的研究

在我国,分类检索语言是众多情报语言中研究历史最长、使用面最广的一种语言,对分类检索语言理论和实践的研究形成了文献分类学。文献分类学的研究内容是情报语言学的组成部分之一。1993 年出版的《中国大百科全书》图书馆学情报学档案学卷和《图书馆学百科全书》都把"文献分类学"作为一个词条予以收入。

面临着新世纪的到来,在本世纪最后一个 10 年中,中国学者们纷纷对分类检索语言的未来及发展趋势进行分析,研究结果认为分类检索语言将朝着分类法主题法一体化、分面组配化、自动化的方向发展。

本文将在后面的章节中专门论述分类法主题法一体化和自动化,这里先概述一下分面组配化的研究情况。

近 10 年来,各种大型综合性分类法,如《中图法》、《科图法》与《人大法》,以及许多专业分类法的修订与编制,都把组配作为提高分类法的适应性、增强分类法生命力的重要措施。关于分类法如何组配化,研究者们提出了许多意见和方法,如提出在等级体系分类法中进一步运用复分和仿分技术及其他组配方式以加强分类法组配功能;也有人提出对现有体系分类法在保留二、三级基本体系基础上进行分面改造;更有甚者提出重新编制一部中国的分面分类法。

特别值得一提的是 1991 年 1 月,《中图法》编委会和全国文献工作标准化技术委员会第五分会在天津召开了"全国《资料法》分面改造学术研讨会",专门探讨了分类法中组配技术应用问题,对

《资料法》分面改造的必要性、原则和方法进行了研究,具体地提出了《资料法》改造的五种模式:(1)采用《国际十进分类法》的模式,扩大组配应用的范围,加强主表和附表的分面分析,采用更多的组配符号;(2)采用《布利斯书目分类法》第二版的基本模式,着重加强主表的分面分析技术,改进标记的合成能力;(3)采用分面叙词表模式,编制分类法主题法一体化的检索语言;(4)采用前苏联《图书、书目分类法》的基本模式,着重改造复分表;(5)综合采用多种组配技术。这些理论研究对我国90年代几部综合性分类法的修订起到了积极的作用。

3.2 分类法的修订和编制

90年代,我国分类检索语言实践主要是完成了对《中图法》、《科图法》和《人大法》的修订工作,使之更加适应分类工作需要。《中国档案分类法》在简本试用的基础上,终于编修完成。此外,在专业分类法领域,也做了一些工作,如《新闻资料分类法》的编制出版等。

(1)《中图法》(第三版)的修订和第四版的出版

《中图法》曾获1985年国家科技成果进步一等奖。该法自1980年刊印第二版以来,经过10年的使用、研究与探讨终于于1990年刊印了第三版。之后,一批热心于《中图法》理论研究和从事实际工作的同行对其展开了从宏观到微观、全面而具体的研究,并取得了一批可喜的成果。研究涉及编制原则、体系结构、编制技术、分类标记、类目设置、类目名称和注释、复分仿分技术及其系列分类表等等。近几年来的研究成果表明分类法的研究正在由原来的纯理论研究朝着实际运用的方向发展,理论研究也逐渐增添了更多的哲学分析和理性思考,这些都为《中图法》第三版的修订工作提供了可靠的理论和实践依据。

1996年7月,《中图法》第五届编委会成立并召开了工作会议,制订了《中图法》第三版的修订方案。其中,对《中图法》第三版的修订原则、修订重点、修订分工进行了重点研究。会议决定,自1996年下半年起,对《中图法》第三版进行修订。修订工作已于1998年12月完成,《中图法》第四版于1999年3月由北京图书馆出版社出版。此次修订确定F经济,TN无线电电子学、电信技术,TP自动化技术、计算机技术为修订重点;区分了类分图书用类目和资料用类目;调整了类目体系,扩充加细了类目,增改了类目注释,如A类增设了“A49 邓小平著作”、“A76 邓小平生平和传记”、“A849 邓小平理论的学习和研究”等类目;增设了指示性类目;增加了沿革注释;修改或规范了类目名称;规范了注释用语和注释引用次序;扩大采用了编列交替分类体系的列类法等。与此同时开展了《中图法》电子版的研究。

此外,中国图书馆图书分类法编辑委员还对《中图法》的系列类表作了相应地修订,如1991年出版发行了《中图法(简本)》(第3版)和《中图法(儿童图书馆、中小学图书馆版)》,1993年又推出了《中图法期刊分类表》(第二版)。

(2)《科图法》第二版的修订和第三版的出版

《科图法》是我国当前通用的三大图书分类法之一。它的第三版于1994年12月由科学出版社出版。这次修订着重增补了有关新学科、新技术、新事物、新问题的类目,适当删除、修改、调整了一些不恰当的类目,酌量增加了一些参见、交替、注释和专类附表。这次修订具有两大特色:一是为各大类增加了必要的文字说明,内容包括大类的学科定义、分类体系和分类规则;二是制订了《中国科学院图书馆图书分类法修改规则》。

(3)《人大法》第五版的修订和第六版的出版

《人大法》于1982年出版了第五版,1989年,中国人民大学图书馆对第五版社会科学类目进行了增订、印刷供内部使用和征求意见,后在此基础上进行全面修订,并于1996年6月刊印了第六

版。这次修订，对社会科学类目，力求详尽地反映学术研究的新成果；对自然科学各类，则根据科学技术的发展和人们认识的深化，做了必要的修订；对具体类目进行修订；对复分表进行了修订；将第十七大类的类名由"综合参考"改为"综合性科学、综合性图书"；进一步明确标记符号的用法；版面由 32 开改为 16 开本，等等。

(4)《中档法》的出版

《中国档案分类法》编制工作开始于 1983 年，1987 年 12 月由档案出版社出版，这是一个简本性质的试用本。经过试用，发现存在着一些问题，如类目比较粗，每个类目类集的档案数量较大，不便检索；有的类目名称不够简洁、准确；个别类目归属不当；注释较少，操作有一定难度。因此，《中档法》编委会从 1989 年起，着手对试用本进行修订，并历时 7 年，于 1997 年由档案出版社正式出版。《中档法》由中华人民共和国档案分类表、新民主主义档案分类表、民国档案分类表和清代档案分类表四部分组成，每一部分又分主表和辅助表。《中档法》的正式出版，对统一档案分类法，实现档案文献分类标准化起到了积极作用。

3.3 编著出版了一批分类学著作

分类学著作是分类学理论化、系统化的成果。近 10 年来出版的主要著作有：中图法编委会编制的《〈中国图书馆图书分类法〉〈中国图书资料分类法〉第三版使用手册》(1991)和《〈中国图书馆图书分类法〉发展方向研究》(1992)，《中图法》编委会主编、安鸿书编著的《〈中国图书馆图书分类法·期刊分类表〉实用指南》(1998)、周继良主编的《图书分类学》(1989 年修订版)，北大图书馆学系《图书分类》编写组编著的《图书分类》(1990 年修订版)，侯汉清、王荣授主编的《图书馆分类工作手册》(1992)，刘湘生主编的《文献分类岗位培训教程》(1993)，陈树年等人的《中国图书馆图书分类法的理论与使用》，刘延章的《文献信息分类学》(1996)和《文献信息分类原则、方法和技巧》，叶千军的《阮冈纳赞理论及其应用研究》(1995)、王金祥、赖伯年、杨邦俊主编的《〈中国图书馆图书分类法〉类名辞典》(1996)，高辉、钟旭著的《〈中国图书馆图书分类法〉〈中国图书馆资料分类法〉(第三版)规范化研究》(1994)，陆宗城主编的《新主题文献分类指南》(1993)，李兴辉、陈树年、戴维民编著的《新学科文献分类手册》(1992)，邓绍兴主编的《档案分类》(1998)，中国档案分类法编委会编辑的《档案分类文集》(1990)等等。另外某些综合性著作也对分类法进行了系统地论述。

4 主题检索语言理论与实践进展

90 年代我国主题检索语言的发展主要体现在叙词语言体系的建立和完善，以及叙词语言新模式和新技术的研究和应用上，展现了一些新特点和新趋势。

4.1 关于叙词语言的理论和技术研究

90 年代以来，我国叙词语言的发展主要呈现出五种趋势：一是一体化词表的编制与应用，二是叙词表编制的计算机化，三是叙词语言的标准化和兼容化，四是传统叙词表的进一步演化、发展及定型，五是叙词表编制的小型化、专业化。本文将在后面章节中分别探讨前三个趋势，这里先概述一下后两种趋势。

90 年代中，传统叙词表在 80 年代的基础上进一步演化、发展与定型。1980 年问世的《汉语主题词表》确立了国内传统叙词表编制的主要原则和具体方法。以后传统词表的编修继承了它以字顺显示为主、系统显示为辅的特点，但又在《汉表》失之繁复的五表俱全模式基础上逐步演化，在追

求功能完备的前提下进行自我优化,实现了词表结构的简洁和实用。此外《军用主题词表》英汉/汉英对照索引的设立打破了汉语叙词表只编英汉索引的惯例,为汉语叙词表的国际化拓宽了道路;《军队档案常用主题词表》的词素式轮排表在国内首次实现了对全部词汇的词素式切分与轮排,增加了标引和检索的入口。这些都是对《汉表》宏观结构的丰富和拓展。

就叙词表的微观结构而言,各项词汇评价指标日趋合理。一般词表的等同率多在10%—20%之间,较好的可达到30%以上。而由于较多地采用等级关系全显示,增加每个叙词款目中参照项,使得关联比和参照度也呈上升趋势。对于先组度,虽然没有明确的界定,但在词表编制中采用组代技术的情况也日渐增多(如《军档表》采用的二元组代),这既防止了一些误组配,又对先组度有所控制。

1996年7月8—9日在北图召开的"全国分类法、叙词表发展学术研讨与成果展示会"上展示了主要是在90年代中新编的近60部叙词表、10余部分类法,其中除了个别几部综合性大型词表以外,都是各个学科、专业方面的叙词表和分类表。这反映出我国词表编制领域,尤其是叙词表编制领域的一个重要趋势:词表编制工作的小型化、专业化。小型专业分类表是对综合大型分类表的补充。而以叙词语言为核心的主题检索系统的发展路向必将是:先着重发展专业文献检索系统,以满足各学科和各行业信息事业发展的要求,然后在此基础上建立综合性的主题检索系统。这是一条符合我国社会需要、读者主题检索需求和主题检索系统发展规律的道路。

4.2 叙词表的编制和修订

90年代,编修了一批传统的和一体化的叙词表。一体化词表,如1994年的《中国分类主题词表》、1996年的《社会科学检索词表》等;传统叙词表,如1991年10月修订的《汉表》和1990年出版的《军用主题词表》及其后编制的系列表。另外还有《国防科学技术叙词表(1991年版)》、1990年9月的《交通专业汉语主题词表》(第四版)和1990年4月编修的《机械工程叙词表(第二版)》,1990年6月的《计算机科学技术汉语叙词表》、1991年10月的《资源科学主题词典》和1995年11月刊印出版的《中国档案主题词表》,等等。

(1)《汉语主题词表》修订出版及其轮排索引的编制

《汉语主题词表》(以下简称《汉表》)是"汉字信息处理系统工程"(简称"748工程")的配套项目,目标是为建立我国计算机情报检索体系奠定基础。该表是由中国科技情报所主编的我国规模最大、收词最多、覆盖面最广的第一部综合性叙词表,于1980年由科技文献出版社出版。但编排不当、词语老化、不易使用等缺点阻碍了它的推广使用。为此,1989年《汉语主题词表》研究小组决定进行修订,且于1991年10月出版。1991年出版的增订本仅限于自然科学部分,将第一版的7个分册,通过版面压缩,合为4个分册:字顺表A-L,字顺表M-Z,词族索引、范畴索引,英汉对照索引。在增订过程中,增补了8221条新词,删除了5434条不适用词,使词表主题词的实用性有了较大的提高;在编排上,按汉字同音、同调、同汉字的顺序排列,使汉字字面成族,符合中国人的查检习惯,方便了检索;订正了原字顺表编辑排版错误及参照关系错误,使词条及词间关系更加准确;增订本关联比达到0.964,参照度达到3.21,都比原本有了很大的提高;注意到了与各专业词表的兼容,提高了该表与国内大多数专业词表的兼容性。

《汉语主题词表》(自然科学增订本)出版后,中国科技信息研究所检索语言研究室和中国索引学会索引技术与索引标准研究室共同为之编制了轮排索引,也由科技文献出版社出版。这是我国目前最大的叙词表轮排索引。它共收录了16万多条索引款目,近600万字。《汉语主题词表轮排索引》按字面成族的原理,较为系统清晰地显示出《汉语主题词表》各款目词之间的等级关系、相关

关系和等同关系，具有《汉语主题词表》字顺索引和一定的范畴索引的功能，弥补了《汉语主题词表》因先组度过高而带来的检索不便的缺陷。在增加入口词的基础上，该索引还可改造成一部质量较高的后控制词表。该索引的编制成功表明了我国轮排索引编制技术具有相当高的技术水平，且已比较成熟。

(2)《军用主题词表》及其系列化建设

《军表》是国家社科"七五"规划基金军事学重点课题，国家军用标准"七五"规划重点项目，由军事科学出版社于1990年12月出版。《军表》共收词52500个，其中正式主题词47340个，非正式主题词5160个，全表由字顺表、范畴表、附表和汉英—英汉对照索引组成，采用计算机辅助编制和等级关系全显示技术，是一部大型综合性军事主题词表，1995年获国家科技进步二等奖。

《军表》出版以后，又编制了《军用主题词表使用手册》、《军用主题词释义词典》、军事文献主题词标引系列规则等配套项目。制定了《军用主题词表编制规则》(GJB1776－93)，开发研制了《军用专业词表微机编制管理系统》，并以此技术为基础编制了军用公文、军用档案、军用机要、军用气象水文、军事测绘与军用地理、军事情报、技侦情报、技侦技术、通信兵、防化兵、防护工程、军用后勤、军用医学、后勤科技、第二炮兵、军用航空、机场工程、国防科技、武器型号、军队政治工作等20部专业词表，基本形成了军用叙词语言体系。

4.3 编著出版了一批主题法研究著作

90年代出版的关于主题法研究的主要著作有：侯汉清等主编的《主题法导论》(1991)，张琪玉、刘湘生主编的《中国分类主题词表教程》(1994)，军用主题词表编制委员会编制的《军用主题词表使用手册》(1990)，邱明斤编著的《主题检索语言》(1990)，张正强编著的《〈中国档案主题词表〉研究与使用精要》(1995)，叶千军、戴维民等编著的《档案主题标引与检索实用教程》(1995)，陈树年主编的《中国分类主题词表标引手册》(1998)，等等。

5 情报检索语言标准化与兼容化进展

5.1 标准化进展

早在80年代，我国就已经对情报语言标准化这一趋势有了清晰的认识，1979年中国情报文献标准化技术委员会的成立并设立专门从事情报检索语言标准化工作的第五分委员会，标志着我国情报语言标准化工作的开始。在80、90年代里，他们制定了下列标准：《汉语叙词表编制规则》(GB13190—91)、《文献多语种叙词表编制规则》(GB15417)、《文献主题标引规则》(GB/T3860—83)、《文献分类标引规则》(报批稿)、《档案分类标引规则》、《同类书排列书次号编制规则》(报批稿)、《文献叙词标引规则》(GB/T3860—95)、《档案文献标引规则》(GB15418)等。

军事文献主题词标引系列规则是1992年国家军用标准计划项目，军事科学院1993年军事科研工作计划课题。它于1994年12月完成，1995年5月国防科工委批准颁布，1995年12月实施，1996年下发全军公文、档案、情报等业务部门使用。由5个分标准组成：《军事文献主题词标引通则》(GTB2418.1)，《军队档案主题词标引规则》(GTB2418.2)，《军队情报主题词标引规则》(GTB2418.3)，《军事图书主题词标引规则》(GTB2418.4)，《军队机关公文主题词标引规则》(GTB2418.5)。这五个标准相互兼容，自成体系，它们的制订和应用为情报检索语言的标准化建设作出了开创性的贡献。

5.2 兼容化进展

情报语言的兼容化是指不同情报语言的语种间对同一文献主题概念的标识可以互换，在先期兼容的情况下，尽可能对主题概念之间的关系采取一致或基本一致的显示。

当今成千上万的数据库和检索系统加入信息互联网络，面对情报语言语种纷繁众多的现实，仅仅依靠标准化已难以应付，而采取兼容化措施则显得更为灵活、有效。所以，当代情报语言的兼容化要比标准化更具重要意义。

通过系列化词表方式达到兼容目的。《中图法》编委会和《军用主题词表》编管会两机构近几年来在这方面作了大量工作，取得了令人瞩目的成绩。如《军表》系列的专业表，目前已达 20 部之多；《中图法》的系列工具书也有十多部。另外国防科技情报系统内的叙词表也自成体系，总量达 11 部之多，在此基础上，由中国国防科技信息中心牵头组建的编制组，对此系列叙词表又进行兼容性处理，并增补新词，于 1997 年编制完成了《电子版国防科学技术叙词表》，并于 1998 年出版了修订版。

分类法主题法一体化词表是我国实现兼容化的另一种主要方式。1994 年我国第一部大型综合性分类表—叙词表对照索引式一体化检索语言——《中国分类主题词表》出版发行。继《教育主题词表》、《农业叙词表》、《音像资料叙词表》等专业性分面叙词表出版之后，1996 年《社会科学检索词表》也编制完成并由社会科学文献出版社出版，它的出现将我国情报检索语言兼容化向前推进了一大步。目前，我国正在形成以《中国分类主题词表》为核心的检索语言兼容体系。

1996－1998 年间我国情报工作者在此基础上又把兼容化研究推向了一个新的高度。他们就实现《中图法》、《科图法》、《人大法》、《汉语主题词表》与国外重要的检索语言如《杜威十进分类法》、《国际十进分类法》、《美国国会图书馆分类法》、《美国国会图书馆标题表》等的对应转换展开了可行性论证。1996 年由山西图书馆与北京图书馆、中国科学院文献情报中心、中国人民大学图书馆共同研制的“计算机文献标引对照系统”便是一部包括了《中图法》、《科图法》、《人大法》和《汉语主题词表》在内的大型兼容检索语言工具，也可说是《中国分类主题词表》的扩展。

6 分类法主题法一体化研究进展

6.1 分类法主题法一体化理论研究进展

分类法主题法一体化语言诞生于 20 世纪 60 年代，80 年代初被介绍到我国。随后，尤其是 1986 年以来，国内研究分类法主题法一体化的论文明显增加，并从实践角度着手开始了一些一体化词表的试编活动（如 1983 年和 1987 年先后编制的《常规武器工业分面叙词表》和《教育分面叙词表》便是最早的尝试），从此一体化词表成为国内情报检索语言研究的又一个热点。90 年代里，我国图书情报界对这一趋势已普遍认同，并在一体化词表的编制中积累了丰富的经验，编制技术已经趋向成熟，陆续编制了 20 部左右一体化词表。这当中包括 3 部大型词表——《中国分类主题词表》、《农业科学叙词表》和《社会科学检索词表》和 7 部中型词表。其中作为国家哲学社会科学“七五”规划重点项目——《中国分类主题词表》和《社会科学检索词表》的编制和出版，对汉语一体化词表的发展起到了推波助澜的作用。一体化词表已经成为一种新颖的、广受欢迎的情报检索语言。这些一体化词表的编制，标志着我国情报检索语言的研究和编制进入了一个新的阶段。

从词表的宏观结构来看，开始于 80 年代末的一体化词表的编制首先是着手于范畴索引和词族索引的合并，后来大胆地将详尽的分类表直接引人叙词表，从而简化了词表的结构，加强了词汇控

制，既便于标引和检索，也便于计算机辅助编表和管理，并提高了兼容性。

从微观结构来看，一体化词表分别为其分类表和叙词表两部分增设了许多构件，同时将类目名称叙词化，并增设了被组代词和聚类词，对输入数据表也进行了相应的加工。这一系列的变化使得语言对文献情报的控制、组织能力大幅度提高，同时标引与检索也更为方便与合理。

经过90年代的理论探讨与实践经验积累，总结出一体化词表共有4种编制模式：(1)改造体系分类表成半分面分类表，然后自动生成字顺叙词表，如《武器工业分面叙词表》就是在《中图法》TJ大类基础上编制而成的；(2)改造传统叙词表，将范畴索引和词族索引改造为分面或半分面分类表，或合而为一组成一个新的范畴索引，如《航空航天技术主题词分面表》和《农业科学叙词表》的编制方式；(3)对现有分类表和叙词表进行对应标引，编制双向对照索引，如《中国分类主题词表》的编制方式；(4)新编分面分类表，再自动生成字顺叙词表，如《社会科学检索词表》和《音像资料叙词表》的编制方式。

6.2 两种类型分类法主题法一体化词表的编制

在我国，分类法主题法一体化词表主要有两大模式：一种是分面叙词表模式，另一种是分类表——主题词表对照索引模式。

90年代里编制的分面叙词表有：1993年出版的《教育主题词表》和《中国人口主题词表》，1994年出版的《农业科学叙词表》，1995年出版的《音像资料叙词表》，1996年出版的《社会科学检索词表》，还有《中国语文教育分面叙词表》和《服装叙词表》等等。

分类表—主题词表对照索引式一体化词表是针对我国分类法和叙词表应用现状的一种一体化检索语言模式。1994年前进行了一些试验性的工作，如1991年编制的《航空航天医学分类主题词表》、1992年出版的《中国图书馆图书分类法(R类)与医学主题词表(MeSH)、中医药学主题词表对应表》(是一种一部体系分类法和两部主题词表的对应表)和1993年出版的《中国图书馆图书分类法教育专业分类表》(只有分类号——主题词对应表一个组成部分)。直到1994年6月《中国分类主题词表》出版才预示着这种一体化语言形式在我国情报语言界地位的确立。

另外，中介词典模式也是值得探索的一种途径。1996年由山西图书馆与北京图书馆、中国科学院文献情报中心、中国人民大学图书馆共同研制的“计算机文献标引对照系统”便是一部包括了《中图法》、《科图法》、《人大法》和《汉语主题词表》在内的大型兼容性检索语言工具，它是《中国分类主题词表》的扩展。

6.3 《中国分类主题词表》和《社会科学检索词表》的编制与出版

(1)《中国分类主题词表》的编制与出版

《中国分类主题词表》在《中图法》编委会领导下，经过全国40个单位160多位专家、学者和图书馆工作者8个寒暑的艰辛劳动编制完成，于1994年6月由华艺出版社出版。它是目前世界上最大的一部一体化词表。它将《中图法》(第三版)与《汉表》有机地结合起来，成为既适用于主题标引，又适用于分类标引的我国第一部大型综合性分类表——叙词表对照索引式一体化检索语言，是我国使用最广的分类检索语言和主题检索语言的兼容互换工具。该词表的问世，把我国一体化词表的研制工作从理论研究阶段推向了推广应用阶段。1996年12月荣获由国家科委、国防科工委、中国科学院、中国科协、国家自然科学基金会五部门联合颁布的“国家优秀科技信息成果”二等奖。

之后，一批专家、学者和图书馆工作者对其进行了大量理论研究和试标引实践，总结出许多经验教训。具体来说有以下一些功能和优点：①由于建立在现有的分类表和叙词表的基础之上，故节

省了编制时间和经费，降低了标引难度，而且能充分利用现有的标引成果；②使分类标引和主题标引结合进行，同时完成，降低了标引难度，提高了标引质量；③实现了两种检索系统的有效互补，从而方便了检索，也提高了检索效率；④可充当分类目录或主题目录的索引，省去其中一种目录的编制劳动；⑤为使用《中图法》的单位补配主题目录提供一条捷径。有利于我国主题标引工作的推广和一体化标引检索工作的开展。

当然，《中国分类主题词表》并非完全兼容的一体化词表，并且它的编制还不完善。通过理论研究和实践分析，发现它还存在以下一些问题：①由于词表的编制是基于先组式学科体系分类法的《中图法》和后组式叙词法的《汉表》，故而这种词表的对应关系必然是不完善的。据统计，其类目的完全对应率为53.4%、基本对应率为17%、大致对应率为6.69%，三者之和为77.09%。②附录的结构与词表的结构性质不符且缺乏必要的注释说明。③还存在一些因编制工程复杂而导致疏忽的问题。

(2)《社会科学检索词表》的编制与出版

《社会科学检索词表》是国家哲学社会科学"七五"重点课题，由中国社会科学院文献信息中心主持研制，朱铁声、吕其苏、侯汉清主编。1993年7月20日至8月10日通过书而鉴定，1996年5月由社会科学文献出版社出版。全表共收词23000个，其中正式叙词18874个，非正式叙词4126个。全表由字顺表、分类表和英汉译名对照表三部分组成，是一部大型的社会科学领域的分类法主题法一体化模式的叙词表。

词表在编制初期，经过研究论证和编制试验，决定采用分而叙词表模式。词表的半分面分类表和字顺表是两个对应互换的组成部分。分类表分17个基本大类，26个一级学科类目，220个二级学科类目。根据社会科学学科特点将学科类目划分为历史、学派、方法论、学科、理论、结构、管理、时间和空间面。因此该表在一体化和分面组配化两个方面均作了有益的探索，并进行了成功的运用。它对词表编制模式、编制技术，直至情报语言学理论都有积极意义。

7 自然语言检索理论与技术进展

自然语言的应用是当今情报检索领域的一种重要的发展趋势。随着计算机在图书、情报、档案、新闻、出版以及其他行业应用的日益普及，自然语言检索也正在我国流行起来，但还只是处于发展初期。值得指出的是自然语言的应用是以计算机检索为前提的，不使用计算机，自然语言检索就难以实现。

7.1 自然语言在情报检索中应用方式

在计算机检索系统中，检索过程都是在程序控制下自动进行的，但标引过程却可以是多种多样的，有许多种标引模式。

(1)无标引方式

包括文本检索、单汉字检索和文本检索作为辅助检索途径。其中，若文本检索的对象扩展到数据库中的全文，则特称为全文检索。目前文本检索是自然语言检索中最普遍的方式。但从检索效率来看，纯粹的文本检索的检全率和检准率都是很不理想。90年代，利用单汉字检索模式研制出了一大批实用化系统。这些系统大致可分为两类：一类是自编中文系统；一类是汉化英文系统。自编中文系统如1990年经济日报社与北方交通大学研制成功的"经济日报新闻资料检索系统"；又如易宝北信信息技术有限公司设计和开发的TRS系统，已经实际运行于人民日报、南方日报、中共中

央对外联络部等国内多家报社、部委机关和信息中心,并成功地开辟了海外市场,在新加坡、台湾投入运行,是目前较为成熟的商业化产品;再如中国化学文献数据库运用支持汉字全文检索的数据结构和检索算法建立起的 CCDOC 系统,在节省存储空间和提高检索速度方面取得了很大进展。汉化英文系统如中国科技情报研究所与 PARALOG 软件公司合作汉化的"全文存贮与管理系统"——TRIP 软件,该软件的汉化 TRIP2.3 版本于 1988 年在 VAX-11/750 机上成功实现了汉字全文检索;又如国家建材局技术情报所和中国科技情报所对国际化的信息检索软件 CDS/ISIS(QUICK IMS)的成功汉化;再如用 LOTUS NOTES 编制的中国名胜诗词库,该单汉字检索系统可实现字符串检索,其单字索引的建立效率很高,100 万字文本作单汉字索引只需要 2 分钟左右。

(2)自动标引方式

包括自动抽词标引、自动赋词标引、自动赋分类号、自动聚类和人机结合抽词标引。自动抽词标引是最早出现的一种自动标引方式,是其他几种方式的基础,而汉语的自动抽词称为自动分词;自动赋分类号和自动聚类是自动分类的两种方式。自动赋词标引需要依靠自动赋词用词表,其实质是在叙词表中大量增添作为入口词的自然语言。自动赋分类号也需要依靠自动赋分类号用分类表,这是词——分类号双向对应表,经由体系分类表分面改造或以分面分类法为基础改造而成,并为分面配备优先次序、分面公式,而且分面优先次序可随需而动。

(3)人工标引方式

包括自由标引和自由词补充标引。自然语言检索本身有许多优点,在时间紧,面又缺少标引力量和充足投资的情况下快速开发数据库新产品,可首先考虑采用自由标引建造自然语言检索系统。1997 年 7 月由解放军报社与深圳科信源电子有限公司等多家单位共同研制完成的《解放军报光盘数据库系统》是我国数据库标引的典范,其分类主题标引在报纸标引中具有开拓性意义,是自然语言与人工受控语言在检索系统中结合的典范。它采用了主题标引、分类标引、人物标引和篇名标引,其中主题标引采用的是自由标引方式,对于这种不使用词表的"非控式"标引,该系统制定了一套针对性较强的详细的标引规则,对标引内容的取舍,标引深度的确定,自由词的措词,副标题的规范,人物、机构名、地名、常见类型新闻和图片新闻的处理都作了具体的规定。在实际的标引过程中,最大限度地使用规范化主题词,提高了标引的一致性。

以上各种自然语言在情报检索中应用的方式,都可借助于后控词表或自然语言与情报检索语言的对应表来弥补其缺点。

7.2 汉语自动分词技术进展

从自动标引中我们可以看出它是以自动抽词为基础的,而对于汉语而言,实现自动抽词首先要解决汉语的自动分词问题。十几年来,对此进行了大量的技术研究,提出了许多分词方案,归纳起来大致有以下几种模式:

(1)词典分词法

是依据抽词词典,辅以其他手段实现分词。如:王永成的部件词典法、北大图书馆学系的主题词表法、毛玉姣等人的关键词表法、陈培久的词典切分组词法。以上分词法都有一定的实用性,尤其是部件词典法,其正确率已达 90%左右,在改进部件词典的情况下正确率还有可能提高,已基本达到实用的水平。

(2)切分标记分词法

包括吴蔚天的非用词后缀表法、周依欣和吴蔚天的链接表法、姚天顺的前后位切割标记表法。

(3)其他分词法

如:邓钦和和龙泽云等人的统计分析分词法、夏海的上下文比较标引法、黄祥喜等的语法语义分析分词法(包括词法分析分词法、句法分析分词法、语义分析分词法)、贺前华和李粟等人的神经网络分词法,还有CWSAIS系统的后控标引法(陈豫等)、词语结构类比标引法(赵宗仁)、汉语文献自动标引专家系统(郭友仁)等。

以上绝大部分分词法都要利用分词词典(即抽词词典),它的缺少正是分词软件普及障碍所在。从实际出发,抽词词典应该走小型化、专业化的道路。

实现文献资源自动标引始终是我国情报界追求的目标,目前研究呈现两种趋向:一是字典法的改进与完善,这类研究多集中于消除歧义分析现象,以提高标引的正确率;二是向智能语义化发展,王永成教授根据文献资源特征又提出了文法章法分析法,具有了一定的实用水平。但影响我国自动标引最终走向实用的瓶颈是切分词典的建立与完善,包括语法知识库的建立。

7.3 后控词表研究进展

任何以词语为文献检索标识的自然语言检索法,都存在检索者构造检索策略困难和检全率较低的问题,并且自然语言检索法也存在着一些影响检准的因素,但若采用后控制措施,这些问题大多数或在很大程度上可以解决。

后控词表的性质类似于入口词表,是一种转换工具,是一种扩检工具,是一种罗列自然语言检索标识供选择的工具。

近10年来,图书情报界对后控词表从理论和实践角度做了一些研究和实验。1988年臧国全对后控词表检索系统进行了研究,并在计算机上建立了一个实验系统。1991年陈源蒸提出了中文图书机读目录主题标引采用后控规范的设想。1993年王源等人在所建立的中国化学文献数据库中建立了以后控规范为基础的标引体系。1994年张琪玉教授在《论后控制词表》(《图书情报杂志》,1994(1))一文中全方位、系统地探讨了后控词表的控制机理、控制程度、编制特点、各种编制方式及其在控制上的差别,提出了一种"分类词表+字顺/轮排表"的结构模式,并说明利用后控词表检索文献的各种方法。同年上海空军政治学院的一个研制小组,在张琪玉提出的总体设计思想基础上研制了《汉语题内关键词索引与后控制词表系统》,其中后控词表的设立,大大提高了对关键词中具有等同语义关系的词的控制,且按分类编排,也使系统既可达到控制同义词(包括准同义词)的目的,又可达到按分类等级体系扩检、缩检的目的。同时,由于收入该后控词表的全部关键词和控制词是经过切分按词素轮排的,因而检索时,查词、选词十分容易,易用性也好。另外陆长旭也系统地提出了后控词表的编制方法。1995年周全明对全文检索系统的后控制技术从理论到实践上作了研究和试验,并设计了一个具体全文后控系统——"军队政工全文检索系统"。北大信息管理学硕士研究生韩冬梅和中科院文献情报中心硕士研究生片玉君也分别完成了题为《后控词表的设计开发与利用》和《后控词表的理论与技术研究》的硕士论文。由此可见,后控词表已成为一个研究热点。然而总体来说,我国对后控制技术和后控词表的研究还只是刚起步,做得还很不够,需要作继续深入地研究。

7.4 自然语言检索中的难题

现行情报检索都可分为标引阶段、检索阶段和标引检索匹配阶段。如果具体细分,标引阶段还可再分为主题分析形成主题概念和将主题概念转换成概念标识两个子阶段;而检索阶段也应包括用户需求分析形成需求概念和将需求概念转换成需求概念标识两个子阶段。从目前自然语言在情报检索中的应用和情报检索的自动化来看,大量的研究还游离于主题分析、用户需求分析和标引检

索匹配三过程之外。而此三点却正是目前自然语言检索的难题所在,牵涉到人工智能中的自然语言理解等重大课题,需要我们对传统观念和思维习惯的突破以及相关科技的应用与突破性进展。

目前此三领域有一些初步研究。如上海交大王永成教授等人对自动文摘的研究就是对主题分析自动化的一种尝试,且已经取得了令人鼓舞的进展;模糊数学在标引检索匹配中的运用是对传统布尔逻辑精确匹配思想的一种突破;我国情报语言学界已经意识到对用户需求分析研究的匮乏和对情报语言易用性研究的忽视,及时洞察到了用户检索行为从"提问——检索"向"浏览——检索"演化趋势。

7.5 自然语言与情报检索语言之争

正当我国对自然语言检索法的研究方兴未艾之际,国内外部分学者提出了"受控词表已经过时"、"情报语言已经进入自然语言时代"、"自然语言必将取代情报检索语言"等观点。针对这一观点,国内外同行们进行了长期的争论。虽然至今仍有部分学者持认同态度,但学术界的主流态度已经趋于明朗:即至少在相当长的一段时间内,自然语言不可能全面取代情报检索语言。由于情报检索过程绝对不能没有控制,而两种语言的优、缺点又是互补的,因而从目前的发展趋势来看,只可能是两者的相互结合甚至融合。

8 情报检索语言计算机化的进展

8.1 词表的计算机辅助编制技术的成熟

计算机辅助编表技术在 90 年代逐渐成熟。据统计,目前我国各系统至少已拥有 100 部以上计算机可读化主题词表。计算机辅助编表技术的进展主要体现在以下几个方面:(1)计算机辅助编表功能拓展。从原先只做些排序、打印等简单编辑工作发展到自动生成词表的组成部分(如字顺表、轮排表等)、进行错误检查、各项统计、以及直接输出词表的各种版本(印刷版、机读版、编微版)等等。(2)计算机辅助编表技术的革新。传统叙词表编制时的计算机输入是按叙词款目填写卡逐条进行的,然后由计算机自动生成反参照、自动校验、自动生成款目、自动排版。到了分面叙词表编制阶段,则采用数据的整体输入方式(输入的数据集,类似于分面分类表,只是附加了一些指示符),这样计算机就可根据输入数据、整体生成词表的各个组成部分,并且基本数据库集中了全部基本数据,冗余度小。如《教育主题词表》和《社会科学检索词表》的编制便采用了这一先进的技术。(3)应用计算机辅助编表技术的主体对象的变化。国内计算机辅助编表起始于 80 年代中后期,但主要限于专业性的中小型叙词表,90 年代起,《军表》将这一技术运用拓宽到了综合性大型词表领域,并研制了《军用主题词表》计算机编表软件。1996 年 5 月,大型综合学科型分类主题一体化分面叙词表——《社会科学检索词表》的出版发行标志着我国计算机编表技术的完善成熟。

8.2 计算机技术在我国情报语言领域应用面的拓展

在我国情报语言领域,计算机技术的应用已经从单纯的计算机辅助编表向词表的编、管、用一体化方向发展。计算机词表管理体现在对词表、类表进行及时地增、删、改,其优点在(1)及时性;(2)词表各部分修改的连动性。计算机词表使用体现在:(1)词表的联机显示与联机标引;(2)自动扩、缩检和改变检索范围;(3)自动统计叙词的标引、检索使用频率;(4)便于用户优化检索策略和为维护词表提供依据;(5)自动提供优选词,等。

在具体的实践活动中，已经研制出了许多编、管结合甚至是编、管、用结合的系统。如《军队档案常用主题词表》课题组在《军表》编表软件的基础上开发出了以词表机读版为核心的标引与检索通用管理软件，包括档案著录、自动标引、联机检索、《军档表》机读版和系统维护五大模块；1994年广东中山图书馆在《中国分类主题词表》机读数据的基础上，研制成功计算机辅助图书分类标引、主题标引与检索系统，并在地区公共图书馆自动化网络中投入实际应用，取得了令人满意的效果；1992年10月《军表》编管会完成了“汉语叙词表微机编制管理系统”的研制，系统包括叙词维护、范畴维护、质量维护、逻辑查错、排版输出和统计分析六大模块以及若干子模块，目前用其编制的和正在编制的词表已有10部左右；1998年版的《电子版国防科学技术叙词表》功能齐全，超越了普通叙词表机读版的功能，而且增挂了叙词导航、定位和叙词范畴选择三个模块来更好地实现经验型计算机辅助标引功能，实现了叙词表编、管、用一体化的模式；又如《军表》编管会研制的“《军用主题词表》应用管理系统”(EMT)于1998年9月25日在军事科学院通过了专家鉴定，该系统设计思想先进，功能结构完备，数据结构科学合理，自动化、集成化、实用化程度高，在整体功能设计、显示浏览方式、自然语言与受控语言自动转换、词表的自维护和自学习能力等方面具有创新性，达到了当前国内外同类技术的领先水平，标志我国叙词语言的研究和应用进入了一个新的发展阶段。1997年7月由解放军报社与深圳科信源电子有限公司等多家单位共同研制完成的《解放军报光盘数据库系统》是我国存储量最大，内容最完整、最集中的军事信息全文数据库，也是我国信息含量最大的电子出版物之一，其分类主题标引在报纸标引中具有开拓性意义，是自然语言与人工受控语言在检索系统中结合的典范。

9 计算机网络环境下的情报语言学研究

联网在我国是一个新生事物，情报检索语言的编制与研究基本上还只是与具体的数据库建设挂钩，关于情报语言在网络中的表现形式和具体应用，国内学术界还没有引起足够重视，只有少量文章谈及此问题，并只是停留在对国外情况的介绍和对国内趋势的预测上。应该指出的是，这是国内甚至世界范围内情报语言学研究的一个前沿课题，对它的研究刻不容缓。以下是国内学术界对其从理论上进行研究的成果。

9.1 网络条件下情报检索语言面临的信息环境的变化

变化主要表现在三个方面：(1)机读数据库迅速崛起，数量急剧膨胀。90年代以来，全世界联机数据库已超过5000个，我国自建的中文数据库也已逾千个。(2)在计算机网络环境中，检索语言的使用对象由原先的专业标引人员和检索人员，扩大到专业不同、层次不一的广大终端用户。(3)数据库类型不断丰富。其中不但包括书目数据库，还有数值数据库、全文数据库、图象数据库等。另外电子出版物也在急剧增长。

9.2 网络信息环境下情报语言的发展动向

可概括为五个动向：(1)网络环境中的情报语言必须具有相当高的兼容性。在多个数据库中检索时，为了方便，也为了提高检索效率，数据库间的兼容是必不可少的，而目前的兼容状况却不容乐观。当前比较有代表性的兼容方案有自动匹配转换、中介词典法、集合词表法、专业词表法、宏观词表法(如傅兰生等进行建立国家叙词库的项目)等。(2)网络环境中的情报语言必须具有相当高的易用性。由于情报语言直接面向终端用户，传统的分类语言和叙词语言存在的标引难度大、速度

慢、词汇更新滞后、对标引和检索人员的专业知识和技术要求过高等弊端就明显地暴露出来了,因而提高情报语言的易用性也就显得刻不容缓了。然而事实上情报语言易用性提高的过程就是情报检索语言不断自然语言化的过程。(3)信息源的无政府状态迫使人们更改判断检索效果的概念和标准。由于信息源分散、无序、更迭和消亡无法预测,用户无法判断网上有多少信息同自己的需求有关,查全率等检索评价指标需要重新定义。浏览、查询、阅览等功能比预期目的的提问检索显得更为实用和迫切,检索语言的联机显示也尤为重要。(4)由于数据源不仅在内容形式上而且在地理位置上都极度分散,因而情报语言对信息的组织功能还应拓展到这一领域。搜索引擎的出现便是为了满足这种需求的,然而目前还没有一个网络检索工具可在检索功能上与传统的计算机检索工具甚至人工检索系统相媲美。搜索引擎是新一代的自然语言检索工具,国内已有几十种,可分为分类目录式和主题索引式,国内分类目录式以"搜狐"为代表,主题索引式以北大的"天网"为代表。但目前的搜索引擎检索机理还很不完善,急待研究和解决。(5)与传统的结构化文本检索不同,网络信息还有图表、图形、图像、声音、动画以及超文本、超媒体等,因此情报检索需要研究多媒体信息特征表现和抽取的索引手段。

10 新世纪的情报检索语言

关于新世纪情报检索语言的研究:较为一致的看法是情报检索语言将进一步得到改进和发展,并将广泛使用。尤其是分类法主题法一体化、计算机化等发展方向更增添情报检索语言的活力。情报检索过程绝对不能没有控制,而且情报检索语言的检索效率在检全率、检准率方面确实优于单纯使用自然语言的系统的检索效率,因此可以预言在今后相当长的一段时期内,情报检索语言将不会被自然语言所取代。

与此同时,自然语言与受控语言结合使用,或者直接采用自然语言检索也是一个重要的发展方向,这方面的研究将越来越多,越来越深入。张琪玉教授就指出自然语言或情报检索语言的未来将是自然语言的情报检索语言化或情报检索语言的自然语言化。"这种新型的情报语言既是情报检索语言(但不是今天的情报检索语言),又是自然语言(但不是现今的自然语言);它既有自然语言的优点而优于今天的情报检索语言,又具有情报检索语言的优点而优于现今的自然语言"。

计算机技术引入情报检索系统,为新型情报语言的诞生提供了条件。如侯汉清教授就认为,随着计算机网络和电子出版物的发展,情报检索语言将会有更多的创新,预计在世纪之交,诸如自动标引、自动分类、自动摘要、自然语言理解、智能情报检索等都会在我们这一代实现。陈树年研究馆员认为体系分类法应向着机读分类法和联机分类法过渡,大力提高其检索功能,向着结构简明、透明度高和使用方法简单的方向发展。张琪玉教授在经过十多年研究和思考之后提出的"学科－事物概念组配型检索语言"是一种新型的理想情报检索语言模式,具有深入研究的价值。另外,随着叙词、自由词、统一分类号在机检系统中的混合使用,带出了整体情报语言的新概念,这种语言以其自然语言与受控语言相结合,分类、主题一体化的全新面貌,代表了情报检索语言发展的新方向。

计算机网络化的普及与深化,对世纪之交的情报语言提出了挑战,同时也为其发展提供了极佳的机会。DDC经历了印刷版到电子版的转变后,其第21版又在国际互联网上发布了网页,提供了网络版。在我国,网络版的检索语言还未出现,对它的研究已经初步开始。曾蕾的《联网环境中情报检索语言的发展趋势与中国未来情报检索语言的研究与发展方向》一文是这方面研究的代表性成果。

可以预见,21世纪必将是我国情报语言学推陈出新、百花齐放、蒸蒸日上的繁荣时期。为了这

一繁荣景象的早日到来，我们应及时地总结过去的经验和教训，充分吸收国内外情报语言学以及其它学科的先进方法和技术，拓宽研究领域，把握时机，建立面向21世纪的情报语言学。

参考文献:

[1]刘湘生.在全国分类法、叙词表发展学术研讨与成果展示会上的总结发言(摘要).中国图书馆学报,1996(6)

[2]周全明.我国情报语言学研究的方向重点及热点.图书与情报,1994(1)

[3]张琪玉.情报语言学基础(增订二版).武汉大学出版社,1997

[4]戴维民.情报检索语言综论.军事谊文出版社,1994

[5]孙勇,王秀玲.我国情报检索语言著作研究.中国图书馆学报,1997(5)

[6]白国应.《中国图书馆图书分类法》的成就、经验、问题、教训与展望.图书情报论坛,1993(3)

[7]白国应.《中国科学院图书馆图书分类法》第三版发行.情报资料工作,1995(2)

[8]白国应.中国文献分类学研究的方向、重点及热点.大学图书馆学报,1995(2)

[9]白国应.走向21世纪的中国文献分类学.中国图书馆学报,1995(3)

[10]白国应.《科图法》的修订原则和修订方法.图书情报工作,1993(3)

[11]白国应.关于修订《人大法》的若干思考.北京高校图书馆,1992(1)

[12]张琪玉.关于文献分类标准化的设想和建议.图书情报论坛,1990(1)

[13]戴维民.中国现代图书分类法.情报检索语言综论,1994

[14]戴维民.《中图法》作为我国标准分类法的几个问题.《情报检索语言综论》,1994

[15]郭宝文.《中国人民大学图书馆图书分类法》第六版修订概况.江苏图书馆学报,1996(4)

[16]雷钧.《中图法》三版修订类目统计分析.图书馆理论与实践,1996(2)

[17]郭茂育,赵水森.《中图法》第三版研究述略.图书馆建设,1996(3)

[18]范亚芳.《中图法》三版统计分析.江苏图书馆学报,1996(2)

[19]陈秀雄.浅谈《中国科学院图书馆图书分类法》第三版的几个特点.图书馆论坛,1996(4)

[20]吴仲强,廖楣彩.中国图书分类法发展趋势论.图书馆论坛,1996(6)

[21]何晓敏.《中图法》修订工作刍议.高校图书馆工作,1996(4)

[22]刘延章,狄九凤.分类法运动在中国.图书馆,1998(3)

[23]叶千军,朱丽.试论80年代以来我国叙词表的发展——兼评《军队档案常用主题词表》.图书情报工作,1997(2)

[24]梅国华等.我国主题检索发展的路向:《汉语主题词表》利用研究.情报理论与实践,1995(5)

[25]张琪玉.情报检索语言的标准化与兼容化.情报理论与实践,1996(5)

[26]张琪玉.叙词表的分面化改造.图书情报论坛,1997(6)

[27]朱芊.《中国分类主题词表》标引分析.北京图书馆馆刊,1997(2)

[28]戴维民.分面叙词表的一次成功实践——《音像资料叙词表》的编制理论与技术.图书情报论坛,1997(2)

[29]张颖.《中国分类主题词表》的研究.图书情报论坛,1995(4)

[30]张琪玉.分类法主题法一体化自动标引系统的基本原理和方法.图书馆论坛,1995(6)

[31]张琪玉.人工语言与自然语言、先控制与后控制的界限在计算机系统中可淡化或取消.图书馆杂志,1997(5)

[32]张琪玉.自然语言与人工语言对应转换——情报检索语言走向自动化之路.中国图书馆学报,1996(1)

[33]张琪玉.论后控制词表.图书情报工作,1994(1)

[34]臧国全.后控词表检索系统研究.情报科学,1988,9(2)

[35]陈源蒸.中文图书机读目录主题标引采用后控规范的设想.中国图书馆学报,1991(4)

[36]王源等.后控规范的计算机处理.现代图书情报技术,1993(2)

[37]陆长旭.后控词表的编制方法.中国图书馆学报,1994(6)

[38]宋明亮.报纸文献机助自由标引研究及对汉语后控制词表动态维护的思考——《解放军报》模拟检索系统设计

实验报告.空军政治学院信息管理系硕士研究生毕业论文集,1995
[39]周全明.全文检索系统后控制技术研究.空军政治学院信息管理系硕士研究生毕业论文集,1995
[40]张琪玉.一个优秀的叙词机助标引系统.图书馆杂志,1999(1)
[41]叶千军.《军队档案常用主题词表》的编制与应用.文献工作研究,1995(2)
[42]张琪玉.探索21世纪的情报检索语言.北京大学学报(信息管理系专刊),1997
[43]张琪玉.学科—事物概念组配型检索语言——关于情报检索语言的遐想与求索.图书馆杂志,1997(2)
[44]赵英莉.联网环境中的情报检索语言研究.图书情报工作,1998(5)
[45]曾蕾.联网环境中情报检索语言的发展趋势与中国未来情报检索语言的研究与发展路向.情报学报,1994.13(5)
[46]刘静.论网络信息检索.图书情报工作,1999(1)
[47]许慧.检索语言的发展及其在网络环境中新型体系的建立.大学图书馆学报,1997,15(2)
[48]马学博.整体情报检索语言与学术刍议.图书情报工作,1995(5)
[49]洪漪.我国情报检索语言学学科建设及其今后的研究课题.情报业务研究,1993.10(1)
[50]侯汉清,陈树年.汉语分类主题一体化词表的进展和技术特色.索引技术和索引标准,1997
[51]吴家柱.汉语主题词表剖析.情报理论与实践,1993(4)
[52]刘瑞.《汉语主题词表》增订本评价.当代图书馆,1995(4)

情报检索研究的认知观与情报检索语言的发展趋势

宁波大学图书馆　胡建平

[摘要]本文阐述了情报检索研究的范式转变:从系统观到认知观,并论述了认知观在情报检索各要素间交互作用的反映,以及情报检索语言的发展趋势。

[关键词]情报检索　认知观　情报检索语言

信息、知识和情报是情报科学的最基本概念,三者之间的关系也是情报科学的重要研究对象。[1]信息是物质的一种普遍属性,是物质存在的方式和运动的规律与特点,它普遍存在于自然界、社会和人类思维活动中。知识是人们在改造世界的实践中所获得认识和经验的总和,它是一种观念形态的东西,只有人们大脑才能产生、识别和利用它。知识是人类特有的信息,是信息的一部分。而"情报就是知识,是知识中动态的、有针对的、有时间性的为未来服务、为人们思考与行动所需要的一部分,是知识中活跃的、积极的、变动的一部分,是知识中的精华,是产生新知识的催化剂。"[2]情报是动态的,信息是静态的,情报不能独立于接受者,它必定与接受者的知识状态有关,一条或相同信息对于不同接受者可能具有不同的情报价值;同样,一条或不同信息对于一个或相同接受者在不同时间也有不同的情报价值,这是因为信息向其接受者所提供的情报取决于他现有的知识状态。由此提出了情报检索研究的认知观。认知观点强调人们的知识状态、信念等媒介着(或交互作用于)他们的情报接收、感知以及生产创造活动。

1　情报检索研究的范式转变:从系统观到认知观

范式是一整套思想,它们表征着本领域的核心观点,充当了本领域系统的知识和研究的一个框架。从70年代末开始至今,情报科学的研究取向又在酝酿另一次重大的转变,即从系统观到认知观。[3]

1.1　情报检索研究的系统观

认知观崛起之前,系统观在情报科学、尤其在情报检索领域占据着主导地位。系统观的正式确立,是以本世纪50年代末在英国的克兰菲尔德(Cranfield)理工学院进行一系列情报检索试验和测试为标志的。[4]这一时期基本确定了后组式索引方式、布尔逻辑检索、倒排文档、查全率、查准率、相关性、匹配等情报检索领域的一些核心概念。实验是这一时代情报检索研究的突出标志,其中关于标引系统与语言的Cranfield实验堪称集大成者,它代表着情报科学尤其是情报检索发展史中的一个真正的里程碑,它的成果一直影响着其后的情报研究的实验方法和设计思想,更是奠定了这一领域的研究对象、原则和方法的框架。它把检索系统视作一个运作中的物理系统,实验环境是一个封闭的物理实验室或是一个变动预先设定的数学模型。研究者在一个模拟情报检索作业的人工或虚拟的环境中进行标引方式、表征手段、检索理论和方法的开发、应用、比较及效果评价。由于明显

引进了物理学或机械工学的研究方法、程序和假说,Cranfield 实验所代表的情报检索研究系统观又被称作情报检索研究的物理范式。这种在历史上对情报研究产生过重大影响的系统观,在 70 年代末受到了挑战,如果我们视情报检索为人(文献著者、标引者、中介者、用户)与人工物(文本、文本表征、索引、检索逻辑与算法、数据库)之间的复杂交互作用,目的是满足人们的现实情报需要的话,系统观的物理范式倾向于关注人工物或记录在物理媒体上的知识的表面表征,而忽略情报系统的用户,经常以系统的角度看待情报利用的主体,要求用户适应系统,研究完全摆脱了人及其认识的主观性因素。在这种背景下,强调理解人的知识结构变化,研究视野开始转向包括人的认知过程与情报行为的更开放的情报检索过程,以不同于物理范式的系统观的立场和方法观察、解释情报现象,人们在情报使用和查询情境中组织与处理情报方式的认知观应运而生了。

1.2 情报检索研究的认知观

认知观点来自于认知科学,认知科学是一个诞生于 50、60 年代的年轻学科,从最初的心理学领域逐渐扩大到涉及哲学、计算机科学、人工智能、社会学、语言学、生物学、人类学等跨学科的综合性研究领域,它以知识为研究对象,由于知识已深人到人类生活的各个层面,研究知识也就成了研究人。

DeMay 指出情报科学认知观的核心思想在于:任何情报的处理,无论是知觉的,还是符号的,都是由构成了情报处理器(如人)的世界模型的整套范畴或概念来传达或协调的。DeMay 对认知观点核心的表述相当明确,得到了情报学领域持认知观点的主要研究者的认同。

Brookes 强调,为了更有效地组织知识(而不仅仅是文献这种载体)人们得考虑主观和客观的认知结构与过程,并提出了基本方程式。Brookes 的弟子 Belkin 基于认知观点,提出著名的“异常知识状态”(Anoomalous State of Knowledge , ASK)假说,[5]将情报学中的认知思想或观点发展成能够解释情报现象,指导情报检索研究的较完整理论。他展开了从认知角度对情报科学的研究,他认为情报科学的基本问题是“促进所需要的情报在情报生产者和情报用户之间的有效传播。”他的“知识非常态”理论认为,之所以有情报需求产生,是因为用户认识到自己存在着知识的非常态以致于无法面对某种问题与情况,而且,用户一般也无法精确描述需要什么来解决这种非常态。因此,情报科学、尤其是情报检索的目的在于描述和解决这种非常态,他的情报科学认知观包括以下内涵:

(1)Belkin 通过“所需的情报”这一术语,把情报科学观念同用户(主体)的观念紧密联系起来,由于把研究的焦点置于“某人需要使用的情报”,这就从观念上有一种从系统驱动到用户驱动的重大变化。

(2)Belkin 提出情报科学同情报生产者和情报用户的内在知识状态有关,为了在情报检索中促进所需要的情报的有效传播,情报科学自身的研究必须关注情报生产者和用户的知识概念结构。

西方另一个情报学家 Vickery[6]认为情报检索主要研究三种知识:公共知识、中介知识和个人知识,公共知识指文献或者说任何记录形式的知识情报;中介知识指分类、文摘或索引等记录知识表征的形式;个人知识指用户的知识状态和当前需求,情报检索研究的系统观倾向于将研究焦点放在中介知识结构上,而忽视公共或个人知识。相反,认知观的情报检索研究关注的是理解和模式化个人知识结构和公共知识,认知观点则尝试识辨和表征存在于公共知识与个人知识中的更深层结构。

2 认知结构在情报检索交互作用中的反映

情报检索是主观的、动态的检索,它以用户的需求为判断准绳,[7]情报检索所得结果的评价须

以用户使用客观效应为前提。情报检索主要由三个因素构成:用户认识到自己存在着知识的非常态,以致于无法面对某种问题与情况,而产生的情报需求;可能包含有用情报的情报源;提供利用情报源手段的检索装置。情报检索的过程,就是上述这三种要素间交互作用的过程。认知观点将情报检索中这三种要素间的交互作用视为不同类型认知结构间的相互作用,认知结构是构成情报处理主体世界表象的范畴和概念体系,在情报检索系统中反映着不同认知来源的生成物或转化物上。认知结构在情报检索不同要素上的反映:

(1)用户的认知空间:工作任务、兴趣、当前的认知状态、以期达到的目标,用户的情报需求和提问涉及反映了不同的认知结构。

(2)情报源的知识结构,著者的文本,包括标题、文摘、图表注释、引文及正文等,代表了著者希望传达的知识结构。

(3)提供利用情报手段的检索装置,包括:系统设计者和生产者、情报检索技术开发者、情报检索语言制订者、标引员及中介者。例:特定的数据库结构、匹配算法和逻辑反映了情报检索系统设计者的认知结构;赋于情报对象的标引词反映了标引员对著者文本内容的解释,因面是标引员认知结构的体现,通常还受到标引词表及标引规则设计者认知结构的影响;自动标引中不同统计、加权算法和规则是其设计者对标引作业的认识及其转化;在需求模式的构造和由中介人(或机器)执行的检索中,把用户的提问转变成检索式的操作也反映了中介人的不同认知结构。

3 情报检索的认知观与情报检索语言的发展趋势

80 年代,认知观已对情报科学产生了巨大的影响,90 年代以来,从认知科学角度研究情报的本质、情报检索等情报科学的问题,已汇集成一种强大的潮流。西方的情报科学研究者们越来越关注情报传递中的人的要素,重视人与信息技术之间的交互过程,情报科学中的认知观在指导和推动情报检索语言的进步方面,也将发挥重大的作用。

3.1 认知观对情报检索语言的影响

情报检索语言是对知识进行组织的工具。知识组织,就是将文献中的主题概念进行有秩序的安排,揭示它们之间的内在联系,形成知识网络,以便于利用。[8]情报检索语言的类型很多,根据情报检索语言的构成原理可分为:以学科为聚类核心的分类检索语言和以事物为聚类核心的主题检索语言;根据情报检索语言的结构可分为:先组式语言和后组式语言,以及体系分类法和组配分类法;根据语言的类别可分为人工语言和自然语言;根据它的标识符号可分为系统序列的号码标识和字顺序列的语词标识。各种语言都具有特定的优点,但都有它的局限性,它们是相互补充的。

情报检索语言是以便于用户利用为目的,它是用户的情报需求通向情报对象的路径。它面对的是用户的不同的认知空间、需求模式、以及用户的情报提问转变为检索式的操作反映的不同的认知结构,因此在用户的认知空间与指向情报对象的不同的认知结构之间的相互作用中,使各种不同的情报检索语言发挥各自的优势和效果,实现相互之间的补充、协调十分重要。情报检索语言的分类表,主题词表,自然语言表征等,都分别反映了对情报对象的不同的认知结构,为用户趋向情报对象提供了可能相同或不同的途径。基于不同认知结构提供的路径进行检索获得和共同情报集合,即不同路径提供的情报集合中的交叠内容或相同项目,与用户需求的相交性最高。

3.2 智能专家系统与情报检索语言

随着计算机技术的发展，人们探求机器模拟人类的智力活动成为可能，人工智能探求人类的思维过程，研究将人类的脑力劳动外延到某种物理装置。而专家系统是人工智能走向实用化的一个最新研究领域。所谓专家系统，是利用存贮在计算机内的某一特定领域内人类专家的知识，来解决过去需要人类专家才能解决的现实问题的计算机系统，它是一种以知识为基础智能化的计算机软件系统。

在智能专家系统中，集中各种检索语言，使"关于情报检索语言的遐想"[9]成为可能。

智能编目专家系统由五部分构成：知识库、综合数据库、推理机、人机接口和知识获取。知识库又分为：综合文献知识库，包括有关文献的各种知识和信息；专业知识库，包括专业领域的概念及概念之间的各种关系；规则库，包括系统控制策略知识和编目专家知识；语言知识库，用来存放编目过程中所需要的语言学方面的知识，主要是字典、词典以及语法知识；用户知识库，包括用户模型知识。

把各种情报检索语言放在专业知识库和规则库。在人机接口处，应用人工智能技术（如知识表示、语言理解、推理和学习等）模拟专家处理问题的认知功能，有效地执行各种认知活动。接受用户的问题时，明白用户的需求，识别用户的认知空间，选用合适的检索语言，产生需求模型，然后把系统对用户的回答、推理结果及对结果的解释等。进行逆转换，以用户可以理解的方式输出给用户。智能专家系统的人机接口处：允许用户以自然的方式表达需求，不受任何人工命令语言或语法的限制，而且能适应用户的不同的认知空间。

各种检索语言都各有所长，又各有所短，不能相互取代，长期以来人们一直在探讨如何使两者结合起来，相互取长补短，从而满足用户多途径检索的需要，在智能专家系统中能够得到实现。

参考文献：

[1]王知津.从情报组织到知识组织.情报学报，1998(3)
[2]肖自力.信息 知识 情报.情报科学，1981(3)
[3]汪冰.试析情报科学研究的若干重点与发展方向.情报科学，1998(4)
[4]邹永利.情报检索研究的理性主义传统与认知观点.中国图书馆学报，1998(3)
[5]Bekm. N. J.. Anomalous state of knowledge for information retrieval. Canadian Journal of information Scieed, 1980 (5):133-143
[6]Vickery, B. C.. Knowledge structures in information reteieval. Journal of Documentation, 48, 1992:326-327
[7]王知津.再论情报的相关性.情报科学，1998(2):71-77
[8,9]张琪玉.学科—事物概念组配型检索语言.图书馆杂志，1997(2):9-12

浅析线性检索语言

福建省图书馆 陈 湘

[摘要]本文分析了线性检索语言的特点:体系结构、数据结构简单实用,严格的受控语言,基于布尔逻辑的条件检索语言,表层性的检索,检索的单一性。探讨论述了线性检索语言的发展趋向,认为在不久的将来,多维检索语言将得到较快的发展。

[关键词]检索语言 线性

1 引言

在过去,由于人们思维表达;信息的记载、存贮、传播、交流与利用方式基本上是线性的特点;人们的信息检索需求属表层性,要求精确、简捷;以及信息数量不大,信息种类不多等原因,线性检索语言一直是人们存贮、检索信息的主要方式,适应了当时社会发展状况,一直到现代电脑技术的出现。由于目前大多数电脑的CPU和存贮方式仍然是线性运行的,因此信息的检索语言至今大部分仍然是线性的,线性检索语言仍然占据着主导的地位。随着电脑技术的发展,检索语言从低级的单维线性向高级的多维方向发展,从线性向非线性的网络、图形等人工智能方向发展。成为下一世纪的主要趋势和主要研究课题。

线性检索语言主要是指以知识的分类表为基础,以严格的受控词,严格的逻辑结构关系排列,并采用符号标示词表关系的检索语言。分类法、卡特词典、主题词法、关键词法等都是典型的线性检索语言,它们以代表知识的符号对知识加以组织、存贮,采用先组式或后组式的方法对信息进行检索,具有简洁、准确、实用的特点。

线性检索语言是在工业化文化背景下发展起来的,它适应了工业化的实用与严谨的文化要求正如许慧所认为的:“等级列举式分类法所反映的是以牛顿理论为代表的传统学科的知识结构,是一种一元决定论的单值线性关系,阐述的是决定者与被决定者的线性关系。”[1]线性检索语言在知识的组织、利用与再生上,发挥了巨大的作用,它使人类自古累积下来的丰富的文化知识,得以很好地组织、存贮与传播。

总之,无论是以科学知识树为基础的分类法,还是主题法,抑或是关键词法等都是人们线性存贮、检索信息的工具,它们提供了快速、简便、结构化的信息检索方式。

2 线性检索语言特点

线性检索语言有着明显的线性特征,是简捷、高效的信息管理系统。

2.1 体系结构、数据结构简单实用

高度规范化、严格地按规定的线性单向排列,体系之间只有从属等级的逻辑关系,没有语义上

的关联,数据结构之间没有描述上的连续性,存贮结构则按顺序排列。

2.2 严格的受控语言

线性检索语言基本上都是经过高度规范化的语言词典,用符号来标示语言之间的关系,符号之间是机械的线性排列,没有任何语义上的连续性和关联性。因此线性检索语言更多的是为图情专业人员所设计,并为他们用于对知识进行描述、揭示、组织与检索,而不利最终网络用户的使用,正如赵文所认为的:“传统检索系统的直接使用者是检索专家。”[2]

2.3 基于布尔逻辑的条件检索语言

线性检索语言之间的语义关联检索和条件检索须要运用到布尔逻辑加以联接,通过布尔逻辑的“或”“非”“否”的单向性选择,将有关语义关系加以比较、联系,从而选出符合检索要求的信息。

由于线性检索语言不能做逻辑乘、除等,因而不能做矩形的检索。此外它不考虑检索过程中实际存在的事物的连续性,对检索的各个环节进行非“此”即“彼”的描述,因此检索是不完全的,输出的结果也没有相关性,此外在描述用户情报总体需求时,缺乏一种全方位的、全视角的描述方法,布尔检索是最基本、最简单的检索模式。[3]

2.4 表层性的检索

线性检索语言提示揭示、描述、组织、检索的是表层性的信息,或文献的集合,不能够深入地揭示和描述检索文献中所涵的信息元,它所检索的只是文献的表面特征,因而不能适应现代信息量的剧增和检索精细化的要求。

2.5 检索的单一性

线性检索语言是基于一维的线性存贮、组织与检索信息的,因而其检索是单一性的,它从一个节点向下一节点逐步地推进,当满足特定条件时,检索结果是按特定类目的文献集,如分类法;当满足一般条件时,检索结果则是一类或几类的文献集,如主题词法。在这过程中错了一步,则下个检索结果即是错误。线性检索语言主要满足了专指性或聚类性的检索。但由于其计算程式简单,因此不能够检索到相关性的信息,不能满足复杂的语义逻辑检索或多样化,多层次的检索。

总之,线性检索语文在揭示、组织、利用知识上,即有其优势,又有短处,它满足了工业化时代文献量剧增的现实以及所带来的知识的有效组织与利用的问题。

3 网络环境下的线性检索语言的发展

网络环境下线性检索语言是否仍有很大的应用范围,这是图情人员所关心的一个问题。在网络环境下,由于信息种类多样化,多媒体化,包括了图表、图形、图像、声音、影视等非文本的信息;在信息结构上有结构化与非结构化之分;以及信息量的剧增;与传统的单纯文本信息已经有了很大的不同;因此社会的信息需求已经有了很大的改变与发展,查寻信息的要求趋于精细化、多样化;“检索用户已发生了根本性变化,不同年龄,不同职业,不同知识前景,不同检索经验的终端用户大量出现”[4]这些都是线性检索语言所无法彻底解决的。网络环境下,人们正在探索、创造更高级、效率更高,能够满足网络社会需求,更好地开发利用信息的检索语言。目前在联机检索中,自然语言检索的研究已经取得了较大的进展,正在逐步取代传统的线性检索语言的主导地位,获得了广泛的社

会应用。

但是，线性检索语言在组织与检索文本信息中仍有较明显的优势，它的实用、速度与效率是非线性检索语言所不及的。目前线性检索语言的研究与完善仍是检索技术研究中的重点，它在网络中仍具有一定的作用。

青年学者洪漪在论文中认为分类法“在网络中可以用于联机浏览检索，非文本信息的组织，超文本系统的管理，并可以作为网络信息组织的通用工具，促进分类主题一体化的发展。”她还描述了应用于信息网络的未来分类法的主要特点，即机读化。国际通用性。兼容性。灵活性等。[5]

目前线性检索语言主要向二个方向发展：(1)电子化、计算机化。据报导当前 UDC 已经出现了电子版，其中部分内容和检索方式已经有了变化和发展，UDC 电子版化及在计算机上的应用，将更有利于 UDC 的发展，适应信息技术的进步；也更有利于图情界组织、利用、传播知识，也有利于编目的标准化。(2)专家系统的应用。为自然语言检索等人工智能检索提供后控词表，或补助工具，为计算机提供有关智能检索接口，为人工智能检索或专家检索系统提供内部控制与反馈，有利于最终用户按个人习惯和方式检索信息。

4　检索语言发展展望

线性检索语言是在工业化大生产的背景下产生的，并在 100 多年中逐步发展到高潮。计算机的发展，人类信息资源的扩展，以及心理认知学、计算机语言学等学科的研究进展，推动人们创造了基于语义的自然检索语言。自然检索语言是一维非线性的检索语言。自然检索语言是按人的语言方式、思维方式和文化习惯对信息作语义关联的检索，具有人工智能性。目前自然检索语言正向模拟人类思维的基于“语境逻辑”的检索技术发展。

此外数理理论在检索研究中的进一步应用，二维的或多维的检索语言开始出现，并显露出其发展的巨大潜力，“目前在美国圣他·巴巴拉加州大学正在开发研究出一个数字图书馆，研究重点即是多维索引”。[6]

当前向量检索技术就是在二维层面对检索的解析，通过对二个不同范畴的关键词在二维空间中的分布与相关度，通过模糊技术，从而确定检索的内容以及检索深度与广度，并作相关性的输出，从而提高检索的智能性与速度，能够在二维空间上对关键词进行必要的比较，提高检索的效果。

并行计算机的出现和多层存贮器的研制，多维空间的向量检索、代数检索、映射检索等都将得到广泛的应用，成为检索语言的主流。现在国外正在开展代数检索及映射检索等方面的技术研究，多维检索语言的研究正是曾受到国外的研究者的重视。可以预见，在不久的将来，多维检索语言将得到较快的发展。

参考文献：

[1]许慧. 检索语言的发展及其在网络环境中新型体系的建立. 大学图书馆学报，1997(2)
[2]赵文. 最终用户研究. 大学图书馆学报，1997(6)
[3]张进. 情报检索系统：从布尔逻辑到向量空间. 中国图书馆学报，1997(6)
[4]朱丽. 自动标引在自然语言提问中的应用. 情报理论与实践，1998(3)
[5]洪漪，梁树柏. 分类法在信息网络中的应用. 情报学报，1998(1)
[6]黄晓斌. 美国数字图书馆研究的发展. 图书馆建设，1996(6)

机读叙词表编制与维护技术的发展和未来

空军政治学院信息管理系　曹　东

[摘要]本文在回顾机读叙词表发表历程的基础上，讨论了机读叙词表编制软件的一般功能与结构以及对它的评价标准，并介绍了国内关于编制机读叙词表时几项关键的技术，最后探讨了机读叙词表的发展趋势。

[关键词]机读叙词表　叙词表编制　叙词表维护

叙词语言是我国情报检索语言的主流，叙词语言的表现形式是叙词表。词表的编制与维护需要大量的人力和物力，手工编制是一种落后的技术，而且速度慢，成本高，质量难以保证；利用计算机的优良性能进行辅助编制，可降低其复杂度，缩短编制时间，提高效率，保证质量。电子化词表正逐渐成为叙词语言的主要表现形式。

1　国外机读叙词表发展概貌

1966年哈蒙得与ARISE公司就设计了叙词表编制和管理的计算机软件，并应用于《NASA叙词表》和《水源叙词表》的编制工作。

80年代，这项技术日臻成熟，《NCC叙词表》、《INSPEC叙词表》和《UNESCO叙词表》等都是机编词表的成功范例。据统计80年代出版的英文叙词表中45.7%备有机读版。下表为80年代备有机读形式的英文叙词表。可以看出80年代中期是英语语种机读叙词表的大发展时期，它们多与数据库联机检索系统相结合，以提高联机检索系统的检索效率。

叙词表名称	年代	使用方式
ABLEDATA叙词表	1985	ABLEDATA数据库
会计师索引标题表	1985	会计师数据库
伦敦计算机情报服务社叙词表	1986	ACOMPILNE数据库
农业综合企业数据库参考手册;叙词表	1986	AGRIBUSINESS U.S.A数据库
农业术语多语种叙词表:英文版	1982	国际农业科学技术情报系统 近期农业研究情报系统
水研究联机叙词表	1986	Aqualine数据库
生物工程商业情报检索指南;词表	1986	生物工程商业情报数据库
生物伦理叙词表	1986	生物伦理数据库
加拿大商业和时事主题标准档	1985	加拿大商业时事数据库
虐待儿童忽视儿童叙词表	1983	虐待儿童忽视儿童数据库
癌症临床研究词表	1986	癌症临床研究数据库
煤叙词表	1985	煤数据库
咖啡联机检索指南及叙词表	1982	咖啡联机检索数据库

叙词表名称	年代	使用方式
计算机数据库叙词表及字典	1983	计算机数据库
能源数据库主题词表	1984	能源数据库
环境数据库用户手册,词表	1984	环境数据库
电力研究所叙词表	1985	电力数据库
医学史联机数据库词汇指南	1984	医学史联机数据库
哈佛商业评论/联机叙词表	1982	哈佛商业评论/联机数据库
卫生宣传与教育叙词表	1984	综合卫生情报数据库
卫生律师数据库检索手册	1986	卫生律师数据库
高血压情报中心索引词表	1984	综合卫生情报数据库
住房和城市发展研究叙词表	1987	住房和城市发展用户联机系统
国际劳工局叙词表	1985	国际劳工文献工作数据库
工业数据源:数据库用户指南及机读叙词表	1982	工业数据源数据库
纽约时报情报数据库叙词表	1981	纽约时报情报数据库
物理学及工程界情报服务处叙词表	1987	DIALOG 联机显示
医学标题表	1987	医学文献分析与检索联机系统等
国家航空与航天局叙词表	1986	航空数据库等
国家糖尿病情报中心叙词表	1985	联合健康情报数据库
人口情报联机检索系统叙词表	1986	人口情报联机检索系统
康复数据库:国家康复情报中心叙词表	1983	康复数据库
机动车工程师学会叙词表:机动车工程学会电子数据库索引词汇使用指南	1985	SDC 联机显示
工程标题表	1983	《工程索引》机读版
体育叙词表	1985	DIALOG 联机显示
情报存贮公司数据库主题指南	1985	地区商业数据库等
老年词汇叙词表	1986	老年联机数据库
综合词汇叙词表 1950-1984	1985	IFI 综合数据库等
心理学索引词叙词表	1985	DIALOG 联机显示
社会学标引词叙词表	1986	DIALOG 联机显示
联合国书目情报系统叙词表	1985	联合国书目情报系统

数据来源:F. W. Lancaster 著,侯汉清,戴维民,陆宝树译《情报检索词汇控制》中附录——国外叙词表选介。

80 年代末,国外机读叙词表编制技术更加成熟,先后出现了几种通用的编制与管理软件,如英国的 TMS(Thesaurus Maintenance System)、德国的 PROTERM－T 和 Index 等,其中英国的 TMS 软件较为出色。在此基础上,1990 年《INTERNATIONAL CLASSIFICATION》发行专刊对这一时期的机读叙词表编制经验进行了总结。

2 我国机读叙词表发展的三个阶段

我国的机读叙词表研究起步较晚,但其研究进展相当迅速,虽然在一定程度上受益于国外先进编制思想和技术的引进,但在实践和理论上,又能根据我国的实际情况和对计算机技术利用的深化而不断拓宽发展道路。

2.1 辅助编表阶段

我国最早利用计算机进行辅助编表的思想始于 70 年代。其指导思想是尽可能地利用计算机

准确快速的特性，辅助完成词表编制的某些工作。它的主要表现形式是：对文字的编辑处理，排序和打印。

2.2 "编、管"相结合的阶段

真正开始研制计算机辅助编制叙词表管理系统是在80年代中期，以数据库技术的引入为主要标志。

传统的手工编表是编制人员为各个款目词逐条制作卡片，然后根据一定的法则将这些卡片排序形成字顺表、分类表和英汉对照表等，再将其誊抄，付印成书。在编制大型词表时，这种方式的投入与所取得的效益是极不相称的。

随着微型计算机在图书情报领域的应用与普及和国外机读叙词表编制思想的介绍，我国研究人员开始把目光转向了机读叙词表这一崭新的领域，并取得了一批丰硕的研究成果。主要有中科院计算机技术研究所从1983年10月开始研制的词表管理系统(TMS系统)，用于编制和管理《计算机科学技术英语词表》;《火炸药专业叙词表》微机管理系统;《教育科学叙词表》、《教育分面叙词表》管理系统;《军用主题词表》编制管理系统。

这一时期机读叙词表管理软件的功能，主要是针对手工编制方式的各种弱点而设计的，并由最初只做些排序、打印等简单编辑工作发展到自动生成词表的某些组成部分(如字顺表、轮排表等)、对词条进行增、删、改等编辑工作、进行逻辑检查、各项统计和各种版本词表的输出等，一次完成词表的编制与维护工作。其优点体现在：及时性和词表各部分修改的连动性。《军表》将这一技术运用拓宽到了综合性大型词表领域，并研制了《军用主题词表》计算机编表软件。

80年代末90年代初，分面叙词表的出现给词表的机编技术带来了一次革新。数据的输入改过去款目词的逐条输入为数据的整体批量性输入方式，它采用的是数据集的输入方法，在原理和形式上都类似于分面分类表，只是附加了一些指示符。这种方法的数据冗余度小，一次可生成词表的各个组成部分。利用这一技术原理及成果编制的叙词表主要有：《中国分类主题词表》、《科技法规信息管理分类叙词表》、《农业科学叙词表》、《社会科学检索词表》、《标准文献主题词表》、《教育主题词表》、《音像资料叙词表》、《中国语文教育分面叙词表》、《服装叙词表》等。特别是大型综合学科型分类主题一体化分面叙词表——《社会科学检索词表》的机编软件功能完善，推广应用早，成为这类词表机编软件的典范，为分面叙词表机编软件的通用性奠定了基础，它的出版发行标志着我国计算机编表技术的成熟。

2.3 "编、管、用"一体化阶段

90年代中后期，伴随着词表数据库建设和联机检索的发展，机读叙词表开始从单纯的计算机辅助编表向词表的编、管、用一体化方向发展。"用"具体体现在：(1)词表的联机显示与联机标引；(2)自动扩、缩检和改变检索范围；(3)自动统计叙词的标引、检索使用频率；(4)便于用户优化检索策略和为维护词表提供依据；(5)自动提供优选词。

我国自行研制的联机多功能一体化词表编制软件主要有：1993年的《中国人口主题词表》，该软件充分展示了一体化词表在实现查找概念和显示语义关系方面的优越性，促进了利用一体化词表的整体效果；1994年《军队档案常用主题词表》课题组在《军表》编表软件的基础上开发出了以词表机读版为核心的标引与检索通用管理软件，包括档案著录、自动标引、联机检索、《军档表》机读版和系统维护五大模块；1994年广东中山图书馆在《中国分类主题词表》机读数据的基础上，研制成功计算机辅助图书分类标引、主题标引与检索系统，并在地区公共图书馆自动化网络中投入实际应

用,取得了令人满意的效果;1998年版的《电子版国防科学技术叙词表》功能齐全,超越了普通叙词表机读版的功能,而且增挂了叙词导航、定位和叙词范畴选择三个模块来更好地实现经验型计算机辅助标引功能,实现了叙词表编、管、用一体化的模式;于1998年9月25日在通过专家鉴定的"《军用主题词表》应用管理系统(EMT)",与"军用叙词表编制管理系统"相得益彰,该系统突出特征集中体现在一个"用"上,在整体功能设计、显示浏览方式、自然语言与受控语言自动转换、词表的自维护和自学习能力等方面具有创新性,达到了当前国内外同类技术的领先水平,标志我国叙词语言的研究和应用进入了一个新的发展阶段。

3 机读叙词表编制一般规则

3.1 机读叙词表的结构与功能

功能决定结构,结构体现功能。

机读叙词表软件的功能主要应包括以下几方面:

(1)词表辅助编制功能。对普通叙词表来讲应解决词表建造时,数据逐条录入方案,自动生成款目词的各种参照关系,输出印刷型词表的各个组成部分,如:主表、范畴表、汉英对照索引表等。对分面叙词表而言,应解决词表建造时数据整体录入方案,自动生成款目词的各种参照关系,输出印刷型分面分类表和字顺表。

(2)词表维护编辑功能。应解决款目词和范畴的增、删、改,以及范畴的整体搬移等问题,对分面叙词表而言,还应包括对分面标记的编辑功能。

(3)统计分析功能。包括用词量统计,范畴(类目)分布统计,词族分布统计,标引用词频率统计,检索用词频度统计等。

(4)逻辑校对功能。主要有核对非法主题词、非正式主题词、属分关系矛盾、非法范畴号、非法类目、款目词间各种参照关系错乱等问题,还应具备查重去重的功能。

(5)联机显示功能。既可单屏显示叙词表的某一组成部分,如范畴表,分面分叙词表的分面分类表,也可单屏多窗口分别显示叙词表的各组成部分,并保持各组成部分间的连动性;对单个款目词的显示,既可显示其部分关系,也可显示其全部关系。

(6)检索功能。既可提供分类浏览检索,又能提供款目词浏览检索;对检索结果中参照关系的浏览检索,即指即检,不必另行输入检索提问;叙词表间各组成部分保持检索的连动性,字顺表中指针当前位置的款目词,即是英汉对照索引中当前位置的款目词,依此类推。

(7)辅助标引功能。根据用户提取的自然语言主题概念,由系统根据标引规则自动完成对主题概念的分析和识别,辅助用户直接或间接将其转换为词表中的正式主题词。在完成自动优选标引词的过程中,如没有合适的标引词可考虑上位标引,靠词标引或增词自由标引方式。辅助标引功能的高低取决于词表入口率的大小,提高词表的入口率可提高辅助标引质量。在机读叙词表编制软件包中嵌入后控制词表,可有效解决辅助标引的难题。

机读叙词表软件包的结构可包括以下7个模块:(1)词表辅助编制模块;(2)词表维护编辑模块;(3)统计分析模块;(4)逻辑校对模块;(5)联机显示模块;(6)检索模块;(7)辅助标引模块。

研究机读叙词表软件的结构与功能的意义在于:我国机读叙词表软件包通用性差,往往各自为政,不利于数据的兼容与交换,很容易造成因重复性建设而带来的人力和物力的浪费。加强机读叙词表软件包标准化研究,可提高编制效率,节约开发经费,便于各表间数据的交流。

3.2 机读叙词表遵循的主要规则

叙词表的编制应遵循国家标准 GB13190-91《汉语叙词表编制规则》、GB15417《文献多语种叙词表编制规则》、GB/T3860-83《文献主题标引规则》。军事文献应在遵循国家军用标准的基础上，兼顾国家标准。目前颁布的军事文献类军用标准主要有：GJB1776《军用主题词表编制规则》、GJB2418.1-95《军事文献主题词标引通则》和 GJB2418.2-95《军队档案主题词标引规则》。

4 机读叙词表软件的评价标准

从我国机读叙词表发展的历程来看，各种叙词表在其编制与维护过程中，往往采用各自开发的软件，不利于推广应用，准确评价机读叙词表编制管理软件的优劣，既可以提高编制质量，更有利于推广应用。机读叙词表软件的评价不能以叙词表本身的评价原则来代替，虽然可以通过叙词表的完备性、关联比、参照度、等同率、互逆性、清晰度、灵活性、先组度、类目规模等指标来反映机读叙词表编制软件的优劣，但对评价机读叙词表编制软件的功能来讲，这些是不够的。以下是在综合国内外各种机读叙词表编制软件的基础上，提出的评价细则。

A. 基本评价标准

A1.计算机运行环境指标。包括：硬件环境兼容性；基本存储空间要求；操作系统，又可分为单用户和多用户版本。

A2.各类信息的完备性。主要有：最新版本信息；出版发行的各版本信息；使用手册；用户手册以及帮助信息。

A3.价格。包括购买单用户或多用户软件包的价格；升级；技术服务；培训。

A4.售后服务。售后服务机构；服务种类：热线咨询以及培训等有关软件的各种信息。

A5.认同程度。可以从安装用户数量和各类评论文章来评断软件的优劣。

A6.工效问题。主要有：用户界面布置的复杂度，它可从信息结构；颜色配置；窗口技术等方面进行评判；对话方式，它可从命令驱动；菜单驱动；命令与菜单混合型；鼠标驱动等操作方式进行评判。

A7.数据完整性。可从数据备份、数据重组、数据覆盖等过程中是否为自动操作方式以及对存储设备的要求进行评判，此外对于特殊用户访问特殊数据库和功能的权限也应作出相应的规定。

B. 软件功能评价标准

B1. 结构定义

B1.1 款目词及款目词相关属性。主要是对各字段、注释和标记符的最大字符长度的规定以及系统分类与分面分析标记符的区分。

B1.2 关系

B1.2.1 单语种叙词表叙词间关系衡量可从关系定义、预先定义关系数量、关系种类、款目词间关系的数量等方面来进行。关系定义可分为预先定义和用户自定义两种关系。关系种类应具备等价、等级、相关三种关系，系统对各种关系的灵活而全面的处理也应作为考虑因素。对款目词间关系的数量可分别从关系种类进行考查。

B1.2.2 多语种叙词表叙词间关系衡量可从以下几方面进行。对不同自然语言间的对照连接主要考虑最大语种数和主要语种与从属语种间关系的处理。对不同索引语言间的对照连接，可从最大索引语言数、索引语言种类以及连接方式等三方面进行考虑。连接方式可分为：中介词典；直

接对译。

B2. 词表建立与维护时的数据输入

B2.1 数据登录。评价因素有:登录方式,可分为批量性登录;逐条登录;登录的便利性主要考查各种隐性关系的自动反生成。

B2.2 编辑修改。评价因素有:词间关系的自动更正;编辑操作的简易性。

B2.3 删除。评价因素有:批量性删除;词间关系的自动删除;删除操作的简易性。

B2.4 一致性控制。评价因素有预先定义的一致性控制操作和是否具有用户自定义的一致性控制。对款目词及其属性的控制可从去除相同款目词、标记符、语种的一致性控制等功能上进行考虑。对叙词间关系的一致性控制主要有:关系的互逆性控制;款目词间关系的唯一性控制;款目词关系的完整性控制和类目词越级关系控制。

B3. 输出

B3.1 联机显示评价因素主要有:检索款目词的方式,具体可分为浏览式和提问式;独立款目词显示,具体可从属性显示和关系显示进行衡量;词表显示,具体形式应包括:按字顺显示、按题内关键词形式显示、按等级显示、按系统显示(按标记符排列)、按图形显示。按字顺显示时应考虑是否具有等级关系全显示方式。关于叙词表交互式问题的评价,可从语义相关款目词间的链接访问;被选叙词的编辑与删除;表中叙词的直接修改与删除等三方面进行考虑。

B3.2 打印机输出的评价因素可参照联机显示的评价因素。但对一些特殊格式和要求的编排,具体可考虑以下几个问题:分栏多语种对照输出;打印格式的编排:分页;标题;正式词与非正式词在排版上的区分以及用户定义输出方式。

B3.3 文件输出。输出格式分为文本文件和特殊格式文件。

B4. 标引与检索

B4.1 标引的评价因素有:主题概念输入方式是否简便,是否可从屏幕直接点选主题概念;标引规则的完备性;标引正确性。

B4.2 检索的评价因素是检索策略的制定和检索速度。检索策略的制定可具体考虑自动继承检索选项、自动检索相关款目词、扩检与缩检三方面因素。

5 机读叙词表编制的几项主要技术问题和对策

5.1 叙词表的构建技术

一部词表的构建方式主要有两种:逐条式构建和整体批量式构建。这里主要介绍的是分面叙词表整体式构建方法。批量式输入方式优于逐条式输入方式,以下为《人口主题词表》批量输入表的一个片断:

```
0         1         2         3         4         5         6         7
1234567890123456789012345678901234567890123456789012345678901234567890
1A                (N)学科|DISCIPLINES
2A1               (N)  哲学|PHILOSOPHY
3A11              (C)    伦理学|ETHICS
                           J关于什么是人的理想品行的哲学和准则的科学
4A112                        道德/观念|MORAL CONCEPTS
                         心理学|PSYCHOLOGY|
3A12              (C)    社会/心理学|SOLIAL,PSYCHOLOGY
```

4A13　　S 社会学
2A2　　宗教学|RELIGIOUS SCIENCE
　　C 哲学
……………
2A3　　社会/科学|SOCIAL SCIENCES
　　D 人文/科学|Human Studies;Humanities
3A31　　社会学|SOCIOLOGY
3A32　　人口学|DEMOGRAPHY
　　D 人口/科学|Population Science
　　C 社会学
4　　(按理论与应用分)
4A321　　理论/人口学 THEORETICAL DEMOGRAPHY
4A322　　应用/人口学 APPLIED DEMOGRAPHY
4　　(按分支学科分)
4A323　　历史/人口学 HISTORICAL DEMOGRAPHY
.
2A4　　自然/科学|NATURAL SCIENCE
2A5　　技术/科学|TECHNICAL SCIENCE;TECHNOLOGY
　　C 自然科学
……………

表中 1－2 列:本类目/叙词在分类表或字顺表中的级别。

4－15 列:本类目的类号。

16－19 列:本叙词与上级词在字顺表中应建立的参照关系标识,共有 4 种:

空白　本叙词与上级词建立属种关系

[C]　本叙词与上级词建立相关关系

[N]　本叙词不与上级词建立关系

[P]　非叙词,只在分类表中起取聚类作用

20 列开始:中外文类目/叙词。1 级类目/叙词从 20 列开始,2 级从 22 列开始,依此类推。中文叙词中间的/为词素分割符,供建立中文轮排索引使用。

22 列开始:其他参照关系。1 级叙词的关系词从第 22 列开始,2 级从 24 列开始,依此类推。主要有:

J　本叙词的中文注释

D　本叙词的同义词,即非叙词

S　本叙词的上级词

C　本叙词的相关词

计算机在识别此文本建立机读数据时,主要有两个关键步骤:一是如何为各级叙词建立由类级和关系标识规定的参照关系。解决方法是:在逐行扫描数据的同时,随机构建出类级和关系标识的一维数组,当前叙词与上级叙词建立关系时,从后向前扫描数组,找出第一个类级比本叙词类级高的记录,并按关系标识建立关系。二是根据第一步所提供的机读数据,建立其他相应参照关系。每一步运行过程中都提供逻辑纠错,以防建立错误关系。

5.2　自然语言接口

自然语言接口是叙词表中自然语言检索与标引入口,性能优良的自然语言接口能较好地实现自然语言与受控语言的转换,提高自然语言与叙词语言的兼容度,提供多途径检索。但目前主要问题是缺乏自然语言与情报检索语言的对应转换词典,包括汉语的对应转换词典和外语与汉语的对

应转换词典。

1998年通过鉴定的“《军用主题词表》应用管理系统”所提供的自然语言接口，是通过对叙词词素的轮排来实现的，并加入了人工智能的原理和方法，在国内处于领先地位。该系统共构造有词素8万余条，近2万对词素对应关系，20万个转换入口和若干转换规则的知识库。在辅助用户完成标引工作时，文献主题概念可全部用自然语言词表达，若表达文献主题概念的自然词与词表中的叙词或入口词一致，则立即转换为叙词，并将叙词标引登录到标引结果字段；若表达文献主题的自然语言词在词表中没有对应叙词或入口词，则采用智能化的词素相似度模式识别技术，即通过分析主题概念与主题词所含词素及位置的相似程度，自动推荐与自然语言语义相同或相近的主题词，通过人工判断，选定合适的叙词进行标引，实现自然语言与词表中受控主题词的自动转换；若所推荐的叙词均不合适，则可进行组配标引或增词标引。

5.3 超文本技术

机读叙词表中的超文本技术很适合“浏览－检索”方式的联机显示。在实现叙词表语义关系联动显示浏览时，不论点击词表中任意索引表的主题词均可实现多表语义联动显示，一次为用户提供所查主题词全方位语义信息，不仅可大大加快标引(或检索)选词速度，并可提高选词的准确性。

“《军用主题词表》应用管理系统”是最早把这项技术应用于实践的系统。《军用主题词表》的机读词表由字顺表、范畴表、词族索引、轮排索引、英汉对照索引、主题词释义词典六部分组成。可按窗口的叠置或平铺功能，将此六部分索引同时显示在屏幕上，鼠标点击任一词表中的款目词，其余五表均作联动检索，并迅速显示其语义关系，在浏览过程中，点击款目词还可将其登录到标引结果字段。

超文本技术不能完全解决等级关系全显示问题，只能以族首词代替。对它的改进意见是：双击款目词，进入全显示窗口，显示该款目词的全部等级关系。

6 机读叙词表发展的主要趋势

6.1 电子化

即拓展机读叙词表的功能，今后各类机读叙词表的编制维护都应向着”编、管、用”一体化靠拢。

6.2 通用化

实施通用化可提高叙词表编制质量，加快开发速度，避免不必要的浪费。通用化相应要求各机读叙词表软件包的市场化。在我国对于分面叙词表机读版的编制可推广《社会科学检索词表》和《人口主题词表》软件包；对于普通叙词表机读版的编制可推广《军用主题词表》编制管理与应用管理软件。

6.3 网络化

目前，情报检索语言的编制与研究基本上还只是与具体的数据库建设挂钩，关于情报语言在网络中的表现形式和具体应用，国内学术界还没有引起足够重视，只有少量文章谈及此问题，并只是停留在对国外情况的介绍和对国内趋势的预测上。机读叙词表如何适应网络检索环境，这是国内甚至世界范围内情报语言学研究的一个前沿型的课题，对它的研究刻不容缓。主要研究课题有：机

读词表上网服务方式，可分为全表上网浏览检索和简表及功能介绍两种方式；机读叙词表嵌入搜索引擎的问题，在搜索引擎中嵌入机读叙词表主要是为了对搜索引擎实施控制，提高其信息处理加工质量。在嵌入时应考虑网络传输速度等制约因素，合理控制词表规模，可在专业搜索引擎中嵌入机读叙词表来做实验。

6.4 叙词语言与自然语言进一步结合

情报检索语言的自然语言化和自然语言的情报检索语言化是未来发展的方向。但叙词语言与自然语言的结合并不是两者简单的叠加，而是要在保持其严格受控的条件下，充分与自然语言和计算机技术相结合。

6.5 增强叙词语言与自然语言的对应转换能力

一种检索语言与自然语言的对应转换能力越强，它的智能化和自动化水平也越高。提高叙词语言与自然语言的对应转换能力可从以下三方面人手。一是大量增加自然语言作为入口词，提高词表的入口率；二是增设自然语言接口或是其他语义转换途径；三是增强叙词语言的自学习功能，及时把用户在标引和检索中使用的新词汇和建立的叙词与自然语言对应转换关系增补进去，丰富词表的知识内容。

参考文献：

[1] 戴维民．计算机在叙词表编制中的应用．情报检索语言综论，军事谊文出版社，1994

[2] 李仁琪，田钜雷．浅谈计算机辅助编制叙词表．情报理论与实践，1992(6)

[3] 方陆明，王彩华．利用微机辅助编制农业叙词表和建立叙词库管理系统刍议．情报学报，1990，11(4)

[4] 贺玲勇．联机一体化词表的原理与实践——概念查找和语义关系显示的实现．现代图书情报技术，1995(2)

[5] 贺玲勇．联机一体化词表的原理与实践——联机一体化词表的计算机辅助生成．现代图书情报技术，1995(3)

[6] 侯汉清，陈树年．汉语分类主题一体化词表的进展和技术特色．索引技术和索引标准，北京图书馆出版社，1997

[7] Jochen Ganzmann. CRITERIA FOR THE EVALUATION OF THESAURUS SOFTWARE. INTERNATIONAL CLASSIFICATION, Vol. 17(1990) No. 3/4

[8] Claus Ritzler. Comparative Study of PC-supported Thesaurus Software. INTERNATIONAL CLASSIFICATION, Vol. 17(1990) No. 3/4

[9] 张琪玉．一个优秀的叙词机助标引系统．图书馆杂志，1999(1)

《军用主题词表》应用管理系统

军用主题词表编管会办公室　赵建华　赵建国
57319部队　杨肖宁　吴　敏

[摘要]《军用主题词表》应用管理系统是在传统文本《军用主题词表》的基础上研制开发的一种适应计算机和网络环境、多功能、智能化的新型检索语言。本文对该系统的研制背景、研制指导思想、主要功能、数据分析和设计、主要技术创新作了较详细的介绍，并结合该系统的研制和应用，对叙词语言的发展创新进行了理论思考。

[关键词]情报检索语言　叙词表　军用主题词表　应用软件

1　引言

《军用主题词表》(以下简称《军表》)，自1991年7月颁发全军使用以来，对规范全军的情报检索语言，促进全军文献数据库的建设，提高全军文献管理的自动化水平发挥了积极的作用。但是，随着现代信息处理技术的飞速发展，全军各类文献数据库的广泛建立和各种信息网络的相继开通，传统文本词表存在的弊端逐步显露。如：标引费时费力，难以满足数据库快速建设的需要；对使用对象专业知识水平要求过高，非专业人员难以使用；难以对词表进行及时的动态维护和管理，词汇控制滞后；不能与网络联接，难以满足网络终端用户的需要；不便与自然语言进行转换，难以实现与自然语言的有机结合等等。全军广大用户迫切希望对传统的文本词表进行改革创新，研制开发适应计算机和网络环境、智能化的新型检索语言工具。有的用户说：希望能开发出一个像“傻瓜相机”那样的“傻瓜词表”。《军表》遇到的问题也正是当前国内外叙词表研究和应用普遍遇到的理论前沿和技术难点问题。解决这一问题对叙词语言发展创新及提高我军和我国文献管理的自动化水平具有重要的实用价值和学术价值。

1995年初，全军军表编管会决定研制开发《军表》应用管理系统(Electronic Military Thesauri Application Management System，简称EMT系统)。该系统研制开发历时3年完成，1998年9月通过专家技术鉴定。鉴定委员会一致认为：EMT系统设计思想先进，功能结构完备，数据结构科学合理，自动化、集成化、实用化程度高，特别是在系统的整体功能设计、显示浏览方式、自然语言与受控语言自动转换、词表的自维护和自学习能力等方面具有创新性，达到了当前国内外技术领先水平，标志我国叙词语言的研究和应用进入了一个新的发展阶段。

2　系统研制指导思想

课题组通过广泛征求《军表》用户单位意见，走访有关专家和一些开发应用机读词表的单位；学习和研究当前国内外有关联网环境下检索语言方面的学术成果和资料，确立了以下指导思想：

2.1 使用对象

系统要以网络终端用户为使用对象,既可满足广大非专业标引人员的一般需要,又可满足专业标引人员的深层次需要。

2.2 研制基础和引用标准

该系统以《军表》及其配套使用工具(《军表使用手册》、《军用主题词释义词典》、《军事文献主题词标引系列规则》)为数据基础研制开发。主要引用标准有:GB13190—91 文献汉语叙词表编制规则、GJB1776 军用主题词表编制规则、GJB2418.1 军事文献主题词标引通则、GB/P13715—92 信息处理现代汉语分词规范。

2.3 系统目标

总体目标:该系统要功能完备、自动化、集成化、实用化程度高,在技术理论上有所创新。具体目标:实现《军表》应用与自身维护管理的有机结合,集词表浏览、查询、标引、检索、使用统计分析等多种功能于一体;实现《军表》与配套使用工具的有机结合,为用户提供《军表》及其配套使用工具的全面服务;实现自然语言与受控语言的有机结合,同时满足用户使用自然语言和检索系统内部使用受控语言两种需求;实现电脑与人脑的有机结合,充分发挥电脑和人脑两方面优势。

2.4 应用方式

该系统既可独立于文献检索系统单独使用,也可以根据需要开发特定的软件接口,挂入文献检索系统使用。既可以单机使用,也可以在网上多机使用。

2.5 运行环境

该系统硬件环境为PC486,软件平台为中文 Windows 95;开发工具为 Visual BASIC 和 XBase 结构数据库。

3 系统主要功能

3.1 显示浏览

可分别显示浏览《军表》的字顺表、范畴表、词族表、释义词典、英汉表和轮排表,也可单屏多窗口同时显示浏览其中几个或全部索引表;同时对各索引表中的每一款目词均可实现表间语义互联显示浏览,一次为用户提供该款目词的全方位语义信息。

3.2 自动查询

可从主题词、范畴号、范畴名、族首词、主题词英译名、轮排词素等不同入口,对各索引表进行自动查询。

3.3 辅助标引

支持用户在标引中使用自然语言,即由用户根据文献主题提取主题概念,由系统根据标引规则

自动完成对主题概念的分析和识别,辅助用户直接或间接将其转换为词表中正式主题词,实现自然语言与受控语言的有机结合和应用。

3.4 统计分析

可对用户标引中使用的主题概念、主题词、自由词及主题概念与主题词对应转换结果进行自动记录和分类统计分析,为《军表》的动态维护管理提供直接依据。

3.5 显示控制

可对词表显示窗口排列形式、字体和字号设置以及各索引表间语义互联显示进行有效控制

3.6 帮助信息

可向用户提供《系统操作说明》、《军表使用手册》、《军事文献主题词标引系列规则》等帮助信息。

3.7 系统的功能结构

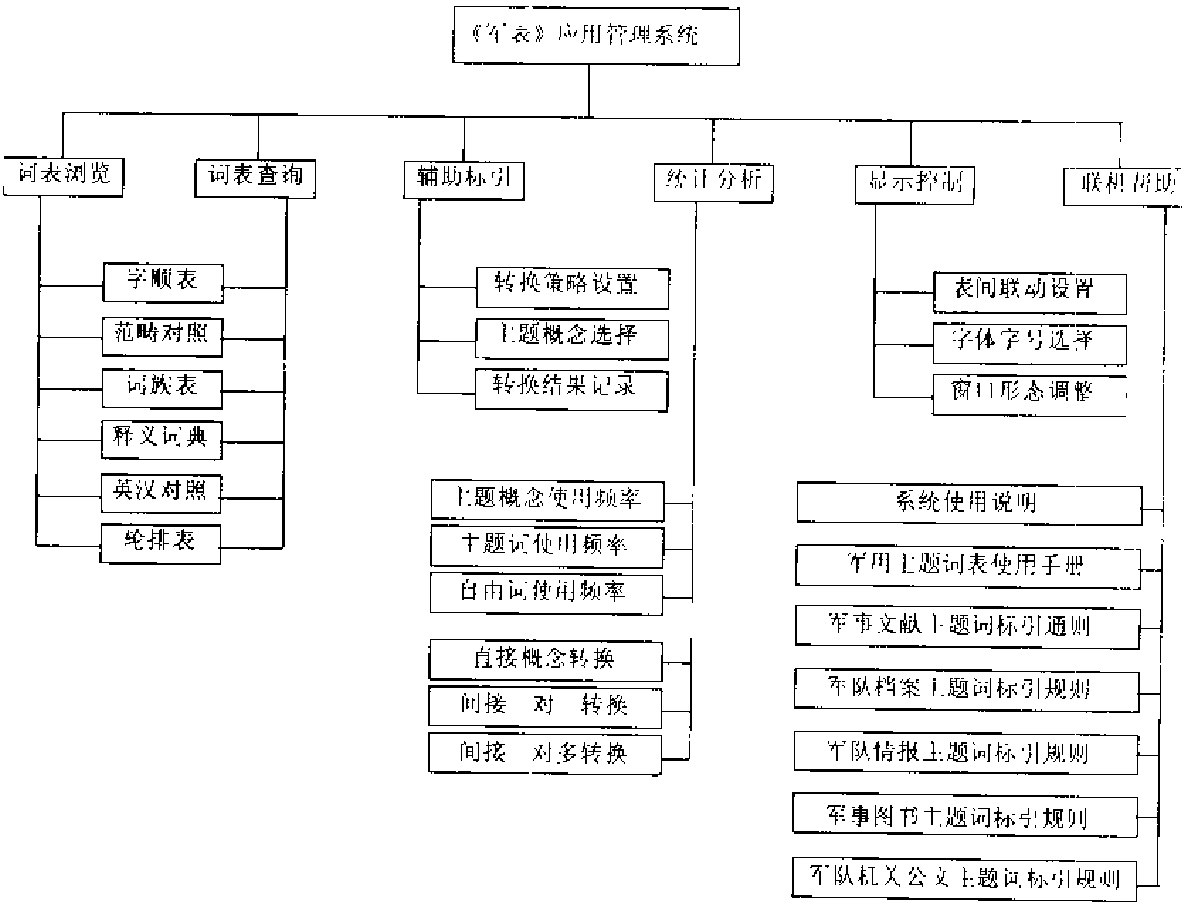


图1 系统功能结构示意

4 系统数据分析及设计

4.1 主题词语义网络结构

根据主题词表编制规则,每个主题词款目的基本数据项有:款目主题词、汉语拼音、英文译名、

范畴代号、注释、范畴名称、用项、代项、属项、分项、参项等。各主题词相互之间构成用代、属分、参照三类关系。各主题词款目之间的不同形式排列，形成一部主题词表的体系结构。

EMT 系统对传统主题词表的体系结构进行了扩充和改革，在主题词款目的基本数据项中增加了释义和词素项。主题词款目的各基本数据项间构成了一个概念语义网络群(见图 2)。

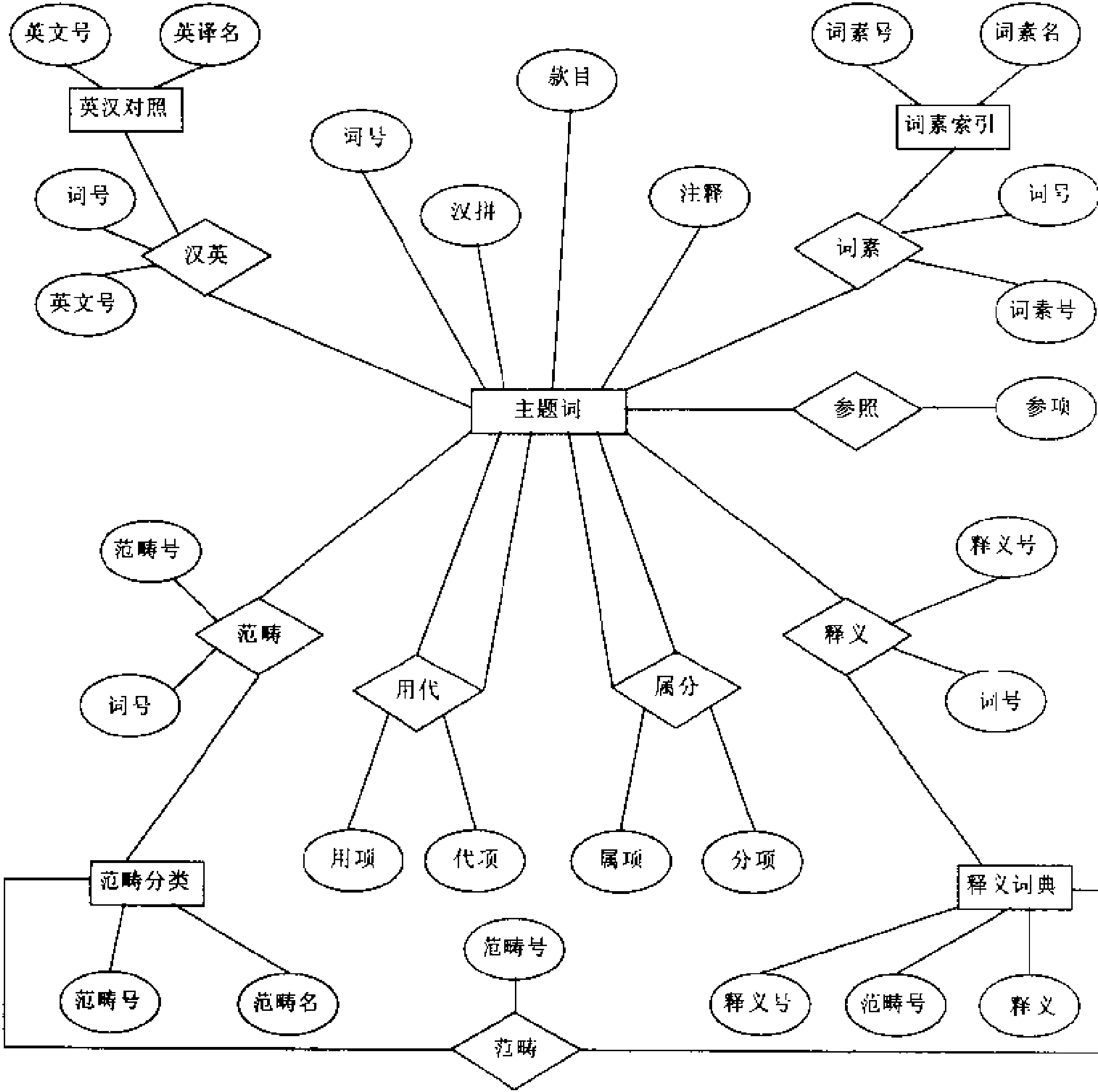


图 2　主题词语义关系结构

4.2　构造转换知识库

为了有效地进行自然语言与受控语言的对应转换，必须构造相应的转换知识库。EMT 系统的转换知识库包括以下数据内容：

· 词素：收录《军表》主题词所含全部词素(表内词素)，以及与《军表》相关的汉语常用词素(表外词素)等，以提高对主题概念的解析能力；

· 词素间关系：收录表内词素之间、表内词素与表外词素之间的同义对照关系，通过增加词素间语义对应关系提高从主题概念到主题词的转换能力；

· 转换规则：主要包括主题概念及其词素标注集合的转换规则；主题词及其词素标注集合的转换规则，主题概念与主题词之间的词素相似度分析转换规则等，以实现从自然语言的主题概念到受控语言的主题词的转换；

EMT 系统基于主题词概念的种属关系采用自完备滚动式分词方案，完成了对《军表》全部主题词的切分整理，并引进同义词林中的相关词素及其对应关系，形成了 8 万多条词素、近 2 万对词素关系、20 多万个转换入口的转换知识库。通过主题词集合与词素集合的映射，以及同义词素之间的相关关系的引入，在不改变词表自身规模的情况下，大大有效地扩展了词表的入口率。同时，为自然语言与受控语言的有机结合和对应转换提供了良好的数据基础。

4.3 系统数据解决方案

EMT 系统的数据存储采用的是基于网状模式的树型层次化的关系数据库解决方案。在确定系统边界及分析信息流程的基础上，利用关系数据规范化理论及超文本构造技术，进行极小化处理，消除数据冗余和关系冗余，增加必要的数据层次及其关联链路，并充分考虑系统存取的响应时间和数据维护的难易程度，形成了合理的适合应用环境的逻辑组织和物理结构。

EMT 系统的全部数据可分为词表相关数据、联机求助数据和图形图像数据三大类。其中词表相关类数据的存储格式采用 XBase 标准；联机求助类数据的存储格式采用 RichTextFormat 标准；图形图像类数据的存储格式采用 BitMap 标准。

· 词表相关数据可细分为基础类数据、词表显示类数据和概念转换类数据三部分。

· 联机求助数据可细分为系统使用帮助类数据、《军表》使用手册类数据和主题词标引规则类数据三部分，由一份操作说明、一本使用手册和五个军用标准构成。

· 图形图像数据主要包括系统主页等信息。

5 系统的主要技术创新

5.1 词表语义关系互联显示浏览

5.1.1 功能设计

EMT 系统采用关系数据库理论和超文本检索技术，使词表的显示浏览方式实现了由传统的线性方式向语义关系互联的多维方式的发展。即用户从词表的任一索引表的任一款目词入手，均可实现与其它各索引语义互联显示，一次为用户提供所查主题词的全方位语义信息。不仅大大加快了标引(或检索)选词的速度，而且提高了选词的准确性。

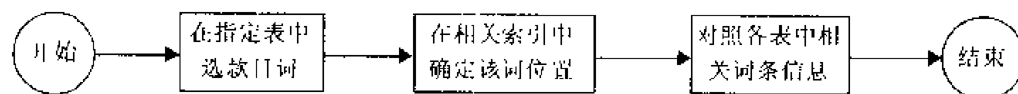


图 3 传统线性浏览词表的基本流程

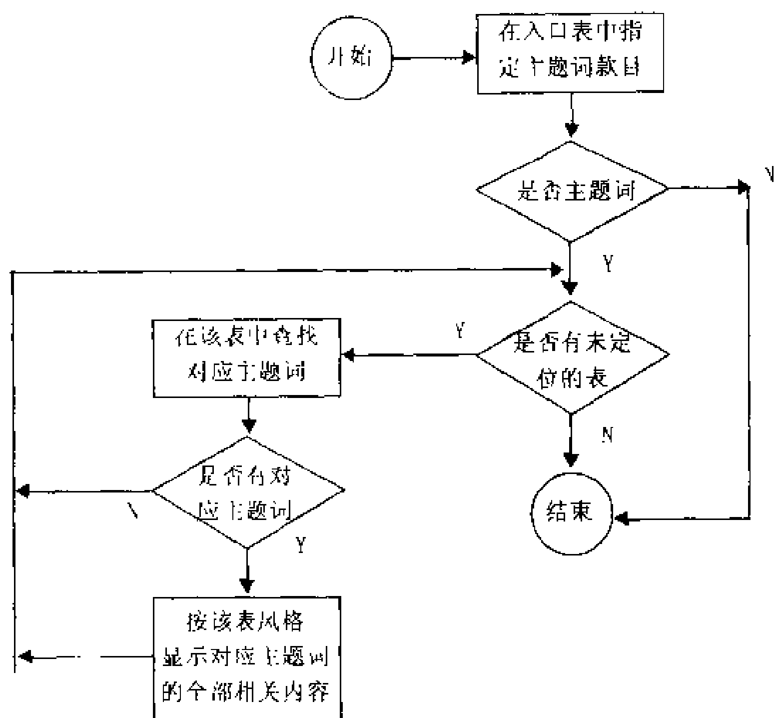


图4 多表语义互联显示浏览基本流程

5.1.2 数据设计

EMT系统根据主题词语义关系结构,各索引表不同的显示格式,通过节点、链的定义将整个《军表》编织成一张语义网,实现了《军表》各索引表间语义互联显示浏览。具体设计是:

· 将每个表的每个款目词均定义为热节点。

· 两两相关节点之间均建立有向索引链进行连接。

· 全部节点由结构链聚合形成整个语义网络。

5.2 自然语言与受控语言对应转换

实现自然语言与受控语言的自动对应转换,是检索语言发展创新的关键环节。目前囿于受计算机技术及人工智能技术发展水平所限,完全由计算机分析文献主题并提炼主题概念,尚不具备条件。EMT系统选择了由人脑进行分析文献主题提炼主题概念的智能化处理,由电脑进行主题概念到主题词的规范化处理,从而完成从自然语言到受控语言的对应转换。为今后实现智能化自动标引做了一次前期探索。

5.2.1 对应转换基本流程

EMT系统具有自然语言入口功能。标引文献时,文献主题概念可全部用自然语言词自由表达。若主题概念与词表中的主题词一致,或与词表中的入口词(同义词和被组代词)一致,都可立即自动转换成主题词,并自动将主题词登录入标引结果字段;若主题概念在词表中没有对应的主题词或入口词,该系统便会对主题概念进行词素分析,利用词素相似度识别转换原理,自动推荐一批含有相同词素的主题词供选择,通过人工判别,选定合适的主题词(包括若干主题词的组配)进行标引;若系统对主题概念不能进行词素分析或所推荐的标引词均不合适,则可将主题概念作为自由词进行标引。

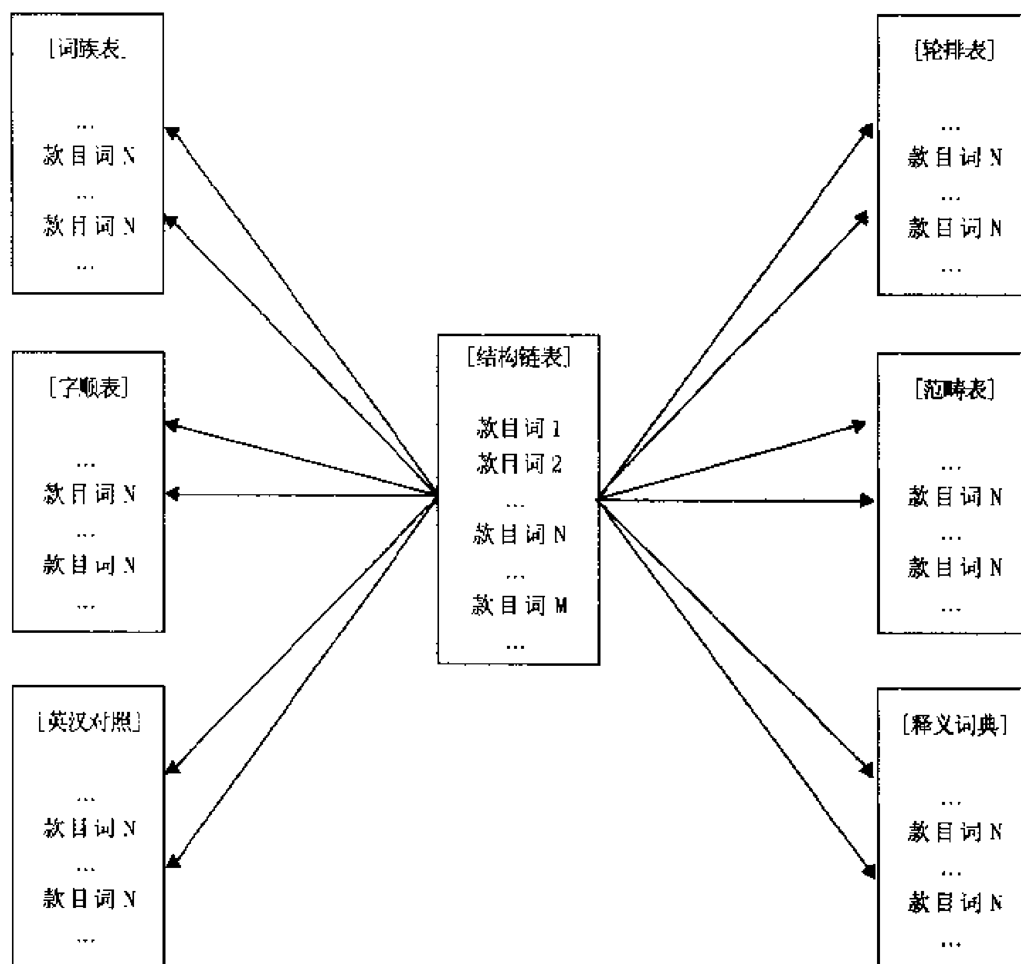


图 5　多表语义互联关系数据结构

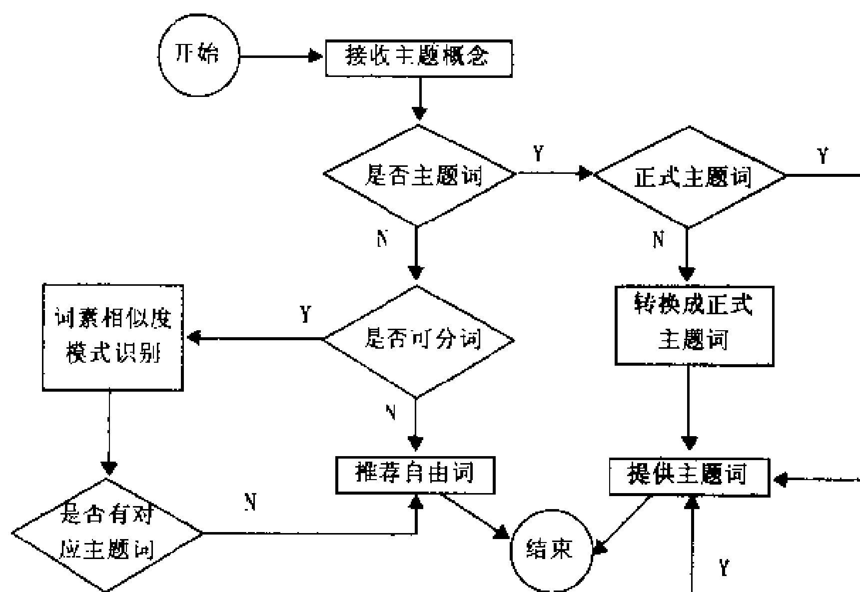


图 6　对应转换基本流程

5.2.2 词素相似度模式识别转换原理

EMT 系统在实现主题概念与主题词对应转换中，对与词表没有对应转换入口的主题概念，采用了词素相似度模式识别转换技术。即通过分析主题概念与主题词所含词素的相似程度，自动选出与主题概念语义相等或相近的主题词供用户选用。

根据汉语词法构造学原理，汉语主题词通常是由多个词素构成。词素可以是单字词或多字词，因此，词表中每一个主题词均可用一组词素及其相关位置进行标注。即，将每个主题词完全映射到一个特定的词素集合上。

同样，用户在标引或检索中使用的每一个主题概念也均可用一组词素及其相关位置进行标注。即，将每个主题概念完全映射到一个特定的词素集合上。通过对主题概念与主题词所含词素相似度分析，便可实现主题概念到主题词的对应转换。

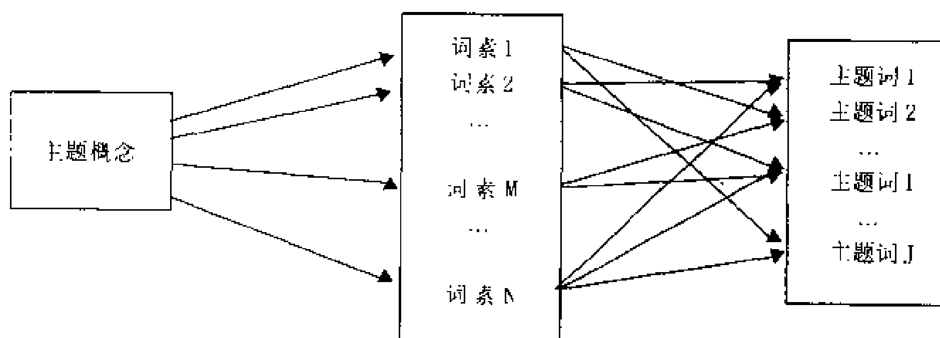


图 7 对应转换原理示意

5.2.3 转换实现

EMT 系统在实现主题概念与主题词的词素相似度转换中，主要采用抽词与赋词相结合的解决方案，具体转换过程：(1)根据转换知识库中的词素及词素同义对照关系，对主题概念进行分词整形，产生对应的词素标注集合；(2)利用转换知识库中主题词的词素标注集合，对主题概念的词素标注集合进行集合运算，形成对应的主题词集合；(3)根据主题概念和主题词所含相同词素的个数、长度及位置特征（词素在词尾、在词首、在词中）进行对应转换计算，自动筛选推荐与主题概念语义相同或相近的主题词，由用户联机验证后产生最终的主题词标引结果。

6. 对叙词语言发展创新的几点思考

通过 EMT 系统研制开发的实践，我们对叙词语言的发展创新产生了以下认识和思考：

(1)随着现代信息处理技术的发展，特别是情报检索的计算机化和网络化，对叙词语言的研究和应用提出了许多新课题、新要求。因此，叙词语言必须适应计算机和网络环境发展创新，否则它将失去自身存在的价值。

(2)叙词语言的发展创新绝非是对其受控原理和方法的否定，也不是简单地把叙词表变为机读形式，而是要在保持其严格受控特色前提下，充分与自然语言和计算机技术结合，克服自身存在的弊端，使之成为适应计算机和网络环境的新型叙词检索语言。根据研制开发 EMT 系统的实践感受，我们认为这种新型的叙词语言的构成原理和方法可概括为：叙词表 + 自然语言 + 自然语言与叙词对应转换 + 计算机技术。

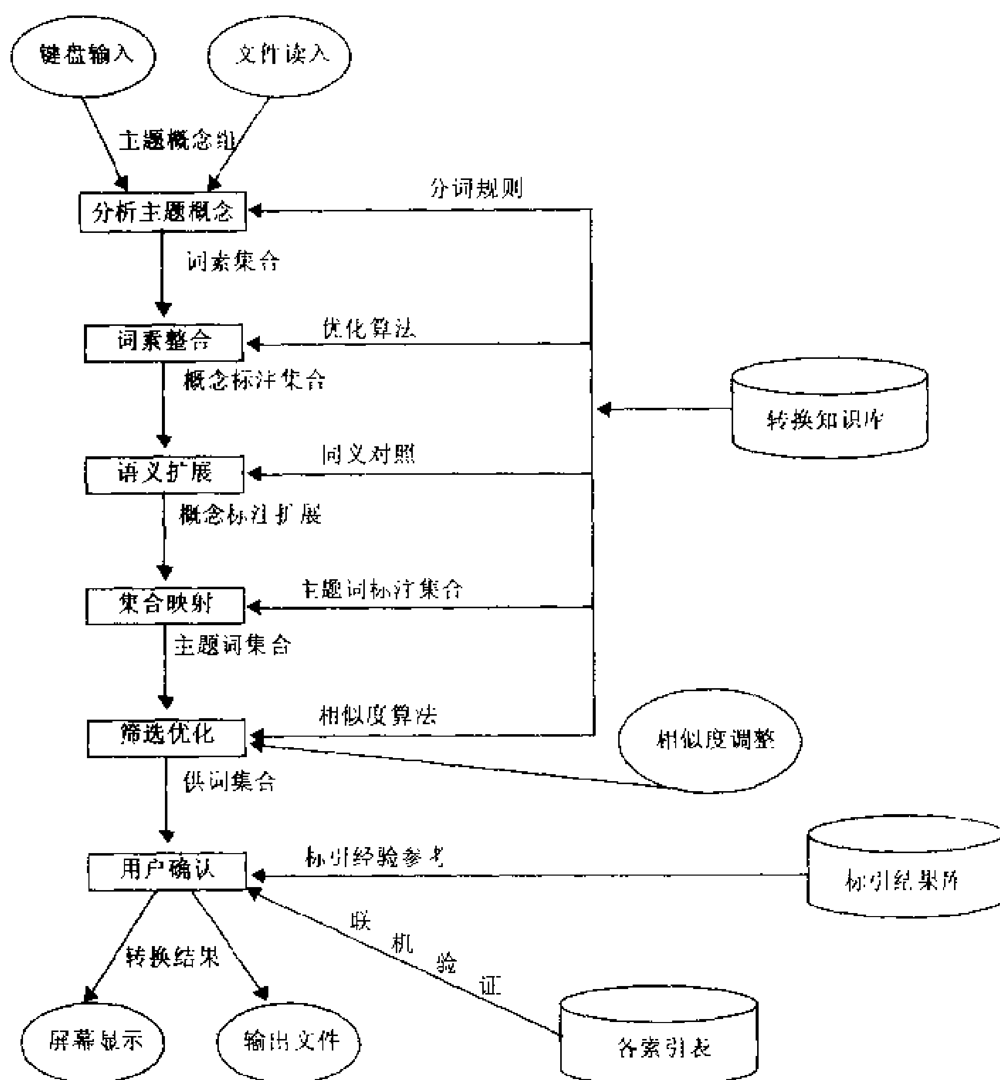


图8　词素相似度模式识别转换流程

(3)增强自然语言与叙词的对应转换能力是叙词语言发展创新的关键环节。根据EMT系统研制开发实践,增强自然语言与叙词的对应转换能力,需做好三个方面工作:一是大量增加自然语言作为入口词,提高词表的入口率。词表的入口率越高,与自然语言的自动转换能力就越强。军表第一版入口率只有10.9%,第二版要增加到50%以上。二是对与词表没有直接转换入口的自然语言,要设法为其提供语义转换途径。EMT系统通过建立转换知识库、分析自然语言与词表中叙词所含词素相似度的方法,实现了自动为自然语言提供语义转换途径,取得了较好的效果。今后随着汉语信息处理技术的发展,特别是汉语自动分词技术的发展,一定会创造出智能化、自动化程度更高的方法。三是要增强叙词语言的自维护能力。EMT系统具有较强的自维护功能,可对用户标引和检索中使用的自然语言概念、叙词、自由词及建立的自然语言与叙词对应转换关系进行自动记录和统计分析,并及时补充到词表中去,从而保证了词表知识内容的不断更新和入口率的不断提高,使自然语言与叙词的对应转换能力越来越强,为叙词语言逐步走向智能化和自动化创造了条件。

参考文献：

[1]张琪玉.情报检索语言的发展趋势(与吴建中的对话).图书馆杂志,1996(4)

[2]张琪玉.自然语言与人工语言的对应转换——情报检索语言走向自动化之路.中国图书馆学报,1996(1)

[3]赵建华.《军用主题词表》的发展现状及展望.情报理论与实践,1996(2)

[4]曾蕾.联网环境中情报检索语言的发展趋势与中国未来情报检索语言的研究与发展路向.情报学报,1994,13(5):325—332

电子版叙词表及其编制技术

中国国防科技信息中心　龚昌明

[摘要]本文概述了自然检索语言和受控检索语言的优缺点，论述了编制电子版叙词表的必要性。结合编制《电子版国防科学技术叙词表》的实践，阐述了编制电子版叙词表的必要条件、编制要求、词表体系结构、建立词间关系的要求、词表显示方式和使用特点。

[关键词]电子版叙词表　词表编制

电子版叙词表，国内外亦称机读词表。它与一般的书本型叙词表有哪些共性、又有哪些特性、如何编制和使用？本文根据《电子版国防科学技术叙词表》编制和使用实践，就电子版叙词表的编制技术，提出一些看法。

1　编制电子版叙词表的必要性

在文献主题标引过程中，必须选择语词将文献主题概念转化为检索标识。在信息检索过程中，也必须选择语词，将用户的信息需求转化检索标识。正确识别和选择语词、实现转化的必要前提是分辩、了解语词的含义。

文献主题标引与信息检索是一个交流过程，存在着语言障碍。如果用户对信息需求的表达主题（检索标识）与文献标引人员对文献主题概念的表达（赋予的标引词）不能沟通，将造成信息检索失败。因此，需要对信息检索和文献主题标引用的词汇进行控制，建立叙词检索语言，把信息检索用户的语言与文献标引人员的语言统一起来。

现可使用的主题检索语言，可分为两种：自然语言和受控语言。自然语言存在多词一义、一词多义现象（若脱离上下文，很难判断其准确含义）、词语的模糊性和不确定性（采用借喻、转义等多种手段表达词汇的含义，往往造成词义含混，难以分辩）、词汇量大、词间关系不明晰等问题。但它具有专指性强、使用方便、费用低等优点。为此，对自然语言进行了控制研究。开发了“后控词表”或“自然语言叙词表”。这种词表，主要由同义词、近义词、准同义词、同根词、语义相关词等组成。但是，汉语自然语言检索尚有不少技术问题有待解决，目前尚无实用的“后控词表”和相应的信息检索系统。

受控语言最明显的优点是对同义词、近义词、准同义词、同形异义词进行了控制，通过识别概念之间的关系，建立起一个与概念体系对应的具有层次结构的术语体系，或称词语体系。在这个体系中，词语之间是相互关联、相互依存、相互制约的。每个词语的含义不再仅由其名称所决定，而主要由它在这个体系中的特定位置来决定。通过这种结构，可以把含义相关的词语有机组织或有机连接起来，以便文献标引人员和信息检索用户选用最能代表某一主题概念的语词。它能满足族性检索、特性检索等多种信息检索要求。因此，国内外大型文献数据库及其检索系统，都采用受控语言。

但是，采用受控检索语言，需要进行复杂的文献主题概念分析，并必须使用受控词表将分析结

果转换成主题检索标识。因而，文献主题标引速度慢，成为数据库建设的瓶颈。此外，由于“在文献标引员的有限专业知识与丰富的文献内容之间，在标引的主观随意性和检索客观要求之间，在死板的词表与灵活新颖的文献用语之间，存在着无法克服的矛盾”，因而，标引一致性差，而且对文献标引人员的要求较高。信息检索用户在信息检索时，必须利用词表，选用最能代表所需主题概念的检索标识或检索入口词。这与他们直接使用熟悉的自然语言相比，肯定要复杂、生疏得多。

此外，在我国文献数据库建设中和卡片式检索系统中使用最普遍的传统的书本型叙词表，存在更新、改版周期长、编制成本高等目前尚难以解决的问题。显然，受控检索语言存在着许多弊端。

随着技术的发展，在现代经济管理与市场信息需求的推动下，各类数据库的建设是确保信息高速公路建设取得社会经济效益的关键。数据库产业是信息产业的基础，又是信息产业的薄弱环节。在我国，从事数据库建设特别是大型文献数据库建设的大多数单位，尤其是其中关键环节——文献主题标引，手段落后，至今大多数依然采用手工作业模式：一人一本词表、一摞工作单、手工填写工作单、人工核对、专职人员录入文摘。致使数据库建设处于效率低、规模小、投资多、见效慢的状态。

文献处理工作如何革新，如何适应现代信息社会对大型文献数据库建设的需要？对此，有不少学者专家进行了多种尝试。对汉语全文检索、汉语文献自动标引、关键词检索等方面进行了大量研究，提出了各种各样的方案，取得了一定的成果。但是，受汉语的特点和计算机人工智能的技术水平等因素的制约，迄今为止，尚不能达到实用程度。

为此，近些年来，国内已有些单位开发了计算机辅助文献主题标引和建库技术，并取得了显著成效。经实际使用表明，在现阶段只有将计算机具有的自动、快速处理数据的功能与人脑的智能结合起来，才是我国当今汉语文献主题标引和文献数据库建设最为切实、有效的途径。

但是，要使计算机辅助主题标引得以实现的前提是必须要有一部机读叙词表，即一部电子版叙词表。因此，当前在我国编制和使用电子版叙词表将是叙词语言发展的趋势和方向。

中国国防科技信息中心于1996年开发了计算机辅助主题标引和建库系统，1997年又对该系统进行了改进，建成经验型计算机辅助主题标引和建库系统。近2年来的使用实践证明，文献主题标引质量和数据库建设的速度有了较大的提高。

可以预料，在汉语全文检索系统和汉语文献自动主题标引研究获得重大突破并投入使用之前，在将大厚本(中、西文)图书资料转换成数字化信息的技术问题取得圆满解决之前，在电子出版物成为图书馆和信息服务部门的重要信息资源之前，计算机辅助主题标引和建库系统将得到推广。由于电子版叙词表具有编制费用低、改版快等优点，随着软件技术的飞速发展和计算机在图书馆、情报资料部门的普及，电子版叙词表必将取代传统的书本型叙词表。可以预料，在我国将出现编制电子版叙词表的高潮。

为支持国防科研和武器装备研制工作，适应“九五”期间国防科技情报事业的发展，实现国防科技情报系统各单位信息的协同处理、数据库联网、终端用户上网检索，1996年，由中国国防科技信息中心牵头，会同有关部委、总公司的科技情报部门，组成联合编制组，历时2年，在国防科技情报系统内使用的11部词表的基础上，进行了兼容性处理，并增补新词，编制成一部《电子版国防科学技术叙词表》(1997年版)。1998年对该部电子版词表进一步增建了用代关系。现版(1998年版)词表收词80005条，其中无关联正式叙词1672条，非叙词22405条，入口率为43%，关联比为97%。该表于1998年装入了中国国防科技信息中心的计算机辅助主题标引和建库系统，使用效果较好。

2 编制电子版叙词表的必要条件

(1)研制和开发编制功能齐全的计算机辅助编表软件系统是编制电子版叙词表的必要条件。该软件必须具有如下基本功能：

(a)录入修改查词功能——能人工将建有词间关系的每个叙词录入系统，人工进行叙词增删改作业；在录入建有非叙词的正式叙词时，能自动生成非叙词；录入时遇到重词，系统自动显示已录入的词及其词号。

(b)词间关系校对功能——能对用代、属分、参照关系及重词、范畴对应进行自动校对或自动纠正，打出错情。

(c)词间关系生成功能——能自动生成族中词库、属项词库、分项词库、族首词库。

(d)词表编排打印功能——能编排打印字顺表、范畴表、英汉对照表、词族表、无关联词表。

在编制《电子版国防科学技术叙词表》过程中，通过不断修改和完善，研制和开发了功能齐全的计算机辅助编表软件系统，为该电子版词表的编成提供了技术保证。

(2)电子版叙词表是供不同水平的文献主题标引人员和终端检索用户使用的，能否快捷地查找到所需的叙词，并尽可能减少键盘操作量，是电子版叙词表能否得到实用的关键。因此，在编制电子版叙词表时，必须同时研制和开发相应的功能完全的软件，为用户提供多种查词手段。

(3)对于在多部差异大、体系结构不同的叙词表基础上编制大型、综合性电子版叙词表，如编制《电子版国防科学技术叙词表》那样，还必须确定合适的编制方法，据此理顺词间关系，将应入选的每个叙词摆到合适的位置，建成科学、正确、完备的语义网。经对 11 部词表体系结构差异的分析，总编组提出采用"比对法"编制《电子版国防科学技术叙词表》：即以《国防科学技术叙词表》(1991年版)为蓝本，其余 10 部词表原收选的叙词，凡经过初审认为应收入电子版叙词表的叙词，应与蓝本一一比对。在比对过程中，对蓝本未收选的叙词，采用以下两种方法建立词间关系：

"接枝法"——凡能纳入蓝本已建词族的叙词，建立词间关系，作为某个或某几个叙词的下位词；

"植入法"——凡不能纳入蓝本已建词族的叙词，确定其词间关系，建立新词族，作为新的族中词，或作为无关联词。

3 电子版叙词表的编制要求

电子版叙词表是供图书馆、科技情报系统各部门信息协同处理和网络终端用户信息检索的动态性叙词检索语言，除应遵循书本型叙词表的编制规则所规定的各项规定外，还应既能在网络环境下运行，即可多机联机辅助主题标引作业，又能单机辅助主题标引作业；非叙词词量，即入口词词量大；查检途径多，方便灵活；遵循叙词表编制的国家标准和国家军用标准；词间关系科学、正确、完备；易于后续更新、增补。

4 电子版叙词表的体系结构

书本型叙词表一般编有主表(字顺表)、索引(包括范畴表、英汉对照索引、汉英对照索引、首汉字拼音索引、首汉字笔画索引、族首词、领词索引)以及附表等。

电子版叙词表不是一般的书本型叙词表简单的机读版，而是一种新颖词表。它的体系结构与一般的书本型叙词表完全不同。通常在机内建有一个主库和几个子库(辅库)，如主库(正式叙词库)、属分关系子库、用代关系子库、参照关系子库、自由词库等。通过专用程序，将主库和各子库有机连成一体，形成一个机读词表。

为使文献主题标引人员利用电子版叙词表进行联机辅助标引，还必须有实时使用模块。这是电子版叙词表的重要构成部分。《电子版国防科学技术叙词表》实时使用模块，由如下3个子模块组成：

(1)叙词导航模块

本子模块为文献标引人员和终端信息检索用户提供多种查词手段，实现将文献主题概念或用户的检索提问转换成规范的叙词。故也可称为"快速查词模块"。

(2)叙词定位模块

本子模块主要实现词族查询和基于经验的叙词联想功能。

(3)叙词范畴选择模块

本子模块主要依据文献主题标引人员给出的叙词，自动将某篇文献归入范畴。

为便于电子版叙词表及时更新和维护，还必须有后期维护模块。本模块主要用于标引(检索)信息、使用过的自由词及其信息的自动保存，词表的动态管理。

5 电子版叙词表的词间关系

(1)电子版叙词表的功能，与一般的书本型叙词表的功能，完全一致，既用于文献主题标引，又用于文献检索(族性检索和特性检索)。因此，在编制电子版叙词表时，必须建立科学、正确、完备的词间关系。词间必须建有属分关系(或上下位关系)、用代关系和相关关系，而不能只建用代关系。虽然，电子版叙词表的体系结构与书本型叙词表不同，但它利用程序根据建立的词间属分关系信息，将全部叙词在纵向构成树形有机体，根据建立的词间相关关系信息，将全部叙词形成横向结构关系。

(2)电子版叙词表不受篇幅的制约，对一个叙词归入的词族及其参照词的数量可不作限制。对词族中下位词的等级，也不作限制。并且可建立大词族，而不采用设置领词的大词族处理技术。《电子版国防科学技术叙词表》(1998年版)建立的词间关系，要比书本型《国防科学技术叙词表》(1991年版)更为完备。词族中下位词的等级从6级提高到10级；从只允许双向成族，改为可按某个叙词所表述的概念的学科属性多向成族。一个叙词是多归入了6个上位词；一个叙词的参照词从规定不大于3个到不作限制，该表中一个叙词的参照词最多达60个。

(3)电子版叙词表是一种机读词表，其可视性比书本型词表差，一次查检所获得的信息量要比一次翻检书本型词表要少得多。这是因为：在编制书本型叙词表时，对"温度计"与"温度表"及其下位词"红外温度计"、"红外温度表"这样的异形同义词，为减少篇幅，一般不作用代处理，有的只对"温度计"与"温度表"作用代处理。其理由是：如用户以首字"温"或首2字"红外"去翻检字顺表，就不难确定词表是否收选了"温度计"、"温度表"、"红外温度计"或"红外温度表"及其词间关系。但对于电子版叙词表，因查词时采用字匹配原理，情况就大不一样。如果电子版叙词表只收选了"温度计"、"红外温度计"，而对它们的异形同义词未作用代处理，则当用户以"温度表"或"红外温度表"去查检时，即便已收选了对应的正式叙词，系统只能告知"叙词表未收选此词"。提高电子版叙词表可视性的最为有效的方法之一是增加非叙词的收选量，即增大人口率，通过计算机自动转换到

对应的正式叙词。只要在文献中出现过的、信息检索用户可能使用的、词形不同的同义词、近义词、准同义词等均应建立用代关系，而不管这个词是族首词或下位词。为此，电子版叙词表的编制人员必须转变观念，把建立词间用代关系、增大入口率作为最为重要的任务，并将其视为电子版叙词表成败、能否实用的关键。

在编制《电子版国防科学技术叙词表》(1997 年版)时，联合编制组虽然对建立用代关系提出了指标要求，但由于尚未根本转变观念，因此，入口率仅达 35%。装入机辅标引系统后，文献标引人员即反映用代关系建立得不够。由是，又从同义词、近义词、反义词、通用名与港台名、全称与简称、学名与俗名、外来语音译名与意译名、英汉混合词与全汉语词、规范汉字与曾用过的简体词、现用名与曾用名、词素相同构词手段或辅助构词手段不同的名词术语等 11 个方面，增建了用代关系。并且，对概念专指度高、词频较低、入选以后会使族中词收词明细度不平衡，并由此引起标引不一致的术语，采用概念组代加以取代。这样，1998 年版《电子版国防科学技术叙词表》的非叙词的收词量已占正式叙词收词量的 43%。现建立的组代关系尚不够，如能根据文献主题标引要求，再增建组代关系，则入口率可进一步提高。

6 电子版叙词表的显示方式

任何一部书本型叙词表主表的显示方式，只能采用一种，或全显示，或上下一级显示，或下分一级上至族首词显示等。通过设计的程序，电子版叙词表可采用多种显示方式。可设计成全显示方式，像书本型叙词表分页显示；也可在键入一个叙词后，在屏幕上显示出词间关系；也可根据要求显示的词间关系等级，显示词间关系。但是，如要求显示二级、三级或全部下位词，或二级、三级或全部上位词时，响应时间会加长。

电子版叙词表的显示方式，一般是由软件设计人员通过设计的专用程序实现的。因此，编表人员应将词表的显示要求，详细地提供给程序设计人员，并要求在使用的算法方面进行优化，以尽可能缩短响应时间。

《电子版国防科学技术叙词表》(1998 年版)装入本中心开发的经验型计算机辅助标引系统后，因优化了算法，尽管它的一个词族的族中词最多有 4700 多个，词间关系的等级达 10 级，显示响应时间仍小于 1 秒。

7 电子版叙词表的使用特点

书本型叙词表可单独出版，独立使用。而电子版叙词表不简单地是书本型叙词表的一种电子出版物，单独出版是没有实际意义的。在采用计算机辅助标引系统的单位，它只是计算机辅助标引系统中的一个必不可少的分系统。在采用文献搜集、著录、标引、检索和目录通报一体化的单位，它只是该种文献处理一体化系统中的一个分系统。在这样的系统中，电子版叙词表作为一个词库，是信息传递载体的集合体。每个叙词(包括对应的英文叙词)是信息传递载体的一个单元。系统在这些单元下集合文献表达的概念信息，从而实现建库、检索、通报等信息活动。由此可见，电子版叙词表只能在计算机辅助标引系统或文献处理一体化系统中使用。既可单机作业，也可网上联机作业。

电子版叙词表是供不同水平的文献标引人员和终端用户使用的，能否快捷地查找到所需的叙词，并尽可能减少键盘操作量，是电子版叙词表能否得到实用的关键。这是由电子版叙词表的使用特点所决定的。故在编制电子版叙词表时，同时要研制和开发相应的软件，为用户提供多种查词手

段。

中国国防科技信息中心研制和开发的经验型计算机辅助标引系统，特设计了一个叙词导航模块，能为用户提供9种查词方式。对此，用户反映较好。

电子版叙词表具有编制费用低、更新改版容易、可在微机或网上使用等优点，可以预料，它必将取代文本型叙词表。在我国，必将兴起编制电子版叙词表的高潮。

参考文献：

[1]侯汉清. 论情报检索中词汇控制. 情报学报，1991,10(4):305-310
[2]苏新宁. 汉语文献自动标引综析. 情报学报，1993,12(4):309-318
[3]谭玉珊. 浅论计算机辅助标引. 情报理论与实践，1996,19(2):24-27

计算机建立分类法和主题词或词串表转换系统的尝试

南京农业大学信息管理系　侯汉清　李　波　戴晶萍

[摘要]本文主要阐述了利用计算机编制分类法和主题词转换系统，即双向对照索引的可行性，并详细说明了该系统软件的设计思想、模块结构和系统功能评价。相信随着该系统的不断完善，它必将取代以往基本上用手工编双向对照索引的方法，促进自动标引、自动分类和联网检索的发展。

[关键词]应用软件　主题词表　图书分类法　检索语言兼容互换

1　机编分类法—主题词或词串表双向对照索引的意义及可行性

60年代中期，西方学者试验研究发现，在分类表和主题词或词串表之间有着非常密切的关系，几乎可以为每个类目找到对应的主题词或主题词串。这一关于分类表和主题词或词串表对应关系的结论导致了一批分类表－主题词或词串表双向对照索引的问世，《中国分类主题词表》(以下简称手工版）就是其中之一。它由国内160位专家、学者和编表人员历经8年的共同努力才告完成，于1994年6月出版，它是一个实现《中图法》与《汉表》之间的兼容互换的工具，目前已成为我国中文图书文献分类和主题标引的主要工具。用户在使用过程中发现，手工版存在着下列一些问题：

(1)其第一卷分类号—主题词或词串对应表中有时分类号对应主题词或词串过少，造成主题词或词串遗漏；有时过多，造成主题词或词串冗余；有时过于粗略，不能较好满足文献标引的实际需要。

(2)手工对应标引，往往以标引员个人的分析猜测为基础，缺乏文献保证，有很大的主观随意性，难以揭示类目实际包含的各种涵义，难以提示类目深层次的隐含主题。

(3)编制过程费工费时，旷日持久，维护更新周期太长。

从分类号和主题词或词串的实质来看，它们都起着为文献类集命名的作用，只是在表现形式上不一样。对字顺主题系统和分类系统的研究也表明：两种系统中任何一种系统只不过是对用同一组织原则进行的另一种组合而已；当这些系统应用于电子计算机时，所有的差别都消失了。因而，解决上述问题的办法之一就是利用现有机读目录(MARC)及期刊论文的书目数据，通过计算机来编制分类法和主题词或词串表的转向系统，即双向对照索引(以下简称机编版)。因为这些数据基本上都包含有分类号(中图法和科图法)及主题词(或词串)等内容，每条书目数据都对应于实际存在的文献，都是标引人员根据他们的知识和经验，并对文献内容进行具体分析而得到的，因而具有广泛的文献保证基础。它大量减少了冗余的主题词或词串，同时增加了一倍以上的必要的主题词及词串；并且其中包含了经过复分、仿分及采用组配而得到的类号和对应词串。从编制情况上来看，持续多年的工程利用计算机编制，仅需几个人在几个月的时间内就可完成，可以节省大量人力、物力和时间。最为重要一点是机编版是现代计算机技术在情报语言学领域应用的产物，它以数据

库形式存在，维护、更新和升级都很方便，可直接与书目数据库联接进行文献检索，并能够实现多种情报检索语言之间的兼容和互换。其在自动标引和自动分类、联机检索方面的广泛应用，必将推动中文文献数据库的建设，促进我国情报检索系统的发展和计算机网络信息资源的开发和利用。

以下是关于如何利用计算机来建立分类法和主题词或词串表转换系统的详细说明。

2 系统用户需求分析

总体上来讲，系统需要完成《中图法》分类号（以下简称中图号）与《汉表》主题词之间兼容互换、科图法分类号与《汉表》主题词兼容互换，中图法与科图法的兼容互换。其中《中图法》分类号与《汉表》的兼容互换是较为典型一个，可代表其它几种情况，因而本文仅以计算机编制《中图法》分类号与对应主题词的标引来作分析。

整个处理过程要求提供 MARC 文件或其它含有中图法和主题词内容的数据库，只需人工简单地进行一些抽取和选择，计算机则可以自动完成中图法分类号与《汉表》主题词的双向对应标引（其结果仍需一些必要的人工判别，删除书目数据中的一些错误标引数据）。

（1）整个系统要求界面美观且便于操作，有关打开或保存文件具有 Windows 格式，可用鼠标选择驱动器、目录和文件名，手工输入文件名时可省略扩展名；存盘时系统自动给出一个唯一的文件名，以供用户选择或修改。

（2）生成的机编版记录应去掉轮排和重复款目，将“－”与“-”，“:”与“：”置换成标准的、统一形式的连字符和组配符。如：

A 农业史:技术史

B 农业史：技术史

C 技术史:农业史

应视为 A 与 B 相同，C 是 A 与 B 的轮排形式，最后仅保留一条记录。

（3）每条书目数据记录操作前应给出一个标识号，以便遇到双类号对应多词串的书目数据时，将类号与主题词串分别搭配生成多个记录，它们标识号保持相同。如：

标识号	第一中图号	第二中图号	第一主题词	第二主题词	…
89	S435.12	S435.11	小麦-病虫害防治方法	水稻-病虫害防治方法	

应生成如下四条记录：

标识号	第一中图号	第二中图号	主题词
89	S435.12	S435.11	小麦-病虫害防治方法
89	S435.12	S435.11	水稻-病虫害防治方法
89	S435.11	S435.12	水稻-病虫害防治方法
89	S435.11	S435.12	小麦-病虫害防治方法

其中第一中图号作为统计频率依据，第二中图号不参与统计，仅供人工分析时参考。

（4）统计结果具有保存（分文本保存和数据库保存两种方式）、浏览、查询、添加、编辑、删除和打印等功能。

（5）最终结果留有接口，可添加补充类名等字段内容。

MARC 文件或外部库文件向指定数据库转换时要保留原来数据所有分类号和主题词；外部数据库转换时只需人工进行简单的选择操作，其它由系统自动完成。

3 设计思想和实现方法

3.1 设计思想

《中图法》分类号与《汉表》主题词之间存在着密切的对应关系，这种关系存在于众多的书目数据库的机读数据之中。通过浏览、统计和分析这种机读数据，可以自动地建立起一个差错率较低的分类法与主题词或词串表之间的转换系统。对机读数据中对应记录的频次统计表明，频次越高，对应记录的有效性就越高。

3.2 开发环境

(1)硬件：主频为 Pentium MMX 166，16 M 内存，1.2G 硬盘，4 速光驱，鼠标。

(2)软件：操作系统为中文 WINDOWS95(OSR2)，程序开发工具为中文版 VISUAL FOX-PRO5.0。

(3)数据：北京图书馆等单位发行的 1975－1997 年的机读目录(MARC)光盘。

3.3 针对用户需求的解决方案

3.3.1 从 MARC 文件到数据库的转换

利用 Foxpro 的低级文件函数功能，采取文件指针移动和字符串处理方式；求出文件中记录总数，让用户选择要操作的记录数据。然后逐一读取“690 ”(中图法号)、“692”(科图法号)或“600－610”(《汉表》主题词)等字段内容，添加到系统指定库中。

3.3.2 接收外部库文件

用下拉列表给出所要转换的库的所有字段名、字段类型和一个对应值，让用户选择与系统库分类号和主题词相对应字段(鼠标点击操作)。利用接收到的选择，将要转换库相应字段名和字段类型修改成与系统库相同的格式，然后添加到系统库中。

3.3.3 对中图号与对应主题词的标引

系统以默认库或用户选择的库中第一中图号前两个为默认统计类号，用户可根据需要修改，但若修改后类号达不到库中记录数的 1/3，系统给予出错提示，继续修改到符合要求为止。

(1)首先将库中主题词字段中所有的组配符号“:”和“：”，“—”和“－”及“-”，规范为统一的“:”、“－”格式。文献标识号的值为库当前记录号的值。

(2)取第二个主题词字段内容，将其从连字符(“－”)处或组配符(:)处分解为若干个词组形式，然后分别与第一、第三、第四和第五个主题词比较。若所有词组均在某一主题词字段中出现，则将此字段置为空。如：

第一主题词	第二主题词	第三主题词
农业－生物环境－环境控制	生物环境－农业－环境控制	环境控制 农业－生物环境

经计算机处理，表示第二、第三主题词为轮排款目，应置空(删除)。

(3)将双类号与多主题词串分别进行搭配，生成多个记录添加在库尾，它们的文献标识号相同，用频率统计和人工判别的方法，筛去错误的对应标引记录。

(4)利用表单、表格等形式显示统计结果；利用 VFP 的扩展类库控件完成查询、编辑、打印等功能。

(5)所有打开的保存文件均采用PUTFILE( )和GETFILE( )函数,它们直接用Windows系统格式打开和保存文件。

(6)在程序执行较长的地方,调用了Visual Foxpro 5.0下的Active X中Microsoft Process Bar Version5.0控件,可根据进程动态显示任务完成情况。

(7)定义每个模块,单击鼠标右键事件为调用帮助系统,设置当前环境下的使用说明。当操作错误或在某些必要地方,系统设置了一些对话框,给用户提示或让用户选择是否继续操作。

4 模块结构和功能

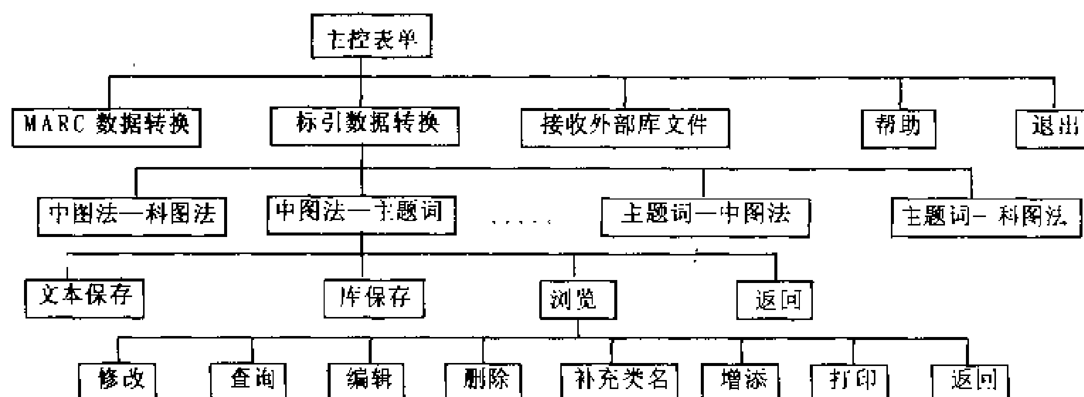
4.1 系统库结构

系统库由7个临时数据表组成,是系统运行时不可缺少的组成部分。其中一个数据表用来临时存贮由MARC或外部库数据转换得到的书目数据,其余6个数据表分别用来临时存贮中图法分类号、科图法分类号、《汉表》主题词三者的对应转换结果。

数据表结构如下表(各个临时数据表根据需要所取字段有多有少,但相同字段名的类型,长度及含义均相同):

字段名	类型	长度	说明
Num	数值型	7	书目数据标识号,取流水顺序号
ztf1	字符型	15	第一个中图法分类号
ztf2	字符型	15	第二个中图法分类号
ktf1	字符型	12	第一个科图法分类号
ktf2	字符型	12	第二个科图法分类号
ztc1	字符型	50	第一个主题词串
ztc2	字符型	50	第二个主题词串
ztc3	字符型	50	第三个主题词串
ztc4	字符型	50	第四个主题词串
ztc5	字符型	50	第五个主题词串
Tj	数值型	4	相同记录的频次

4.2 模块结构



4.3 主要模块功能说明

(1)主控表单

它是系统的初始表单,包含了"MARC 数据转换"、"外部库文件接口"、"退出"及"帮助"等按钮。

(2)MARC 数据转换

实现 MARC 数据文件(纯文本文件)到 VFP 数据库文件的转换,并可保存转换结果。

(3)外部库文件接口

提供已有 Fox 系列数据库文件数据进入本系统的接口。

(4)中图法分类号(科图法分类号)与《汉表》主题词之间的对应转换

本表单中包含有"中图法—科图法"、"科图法—中图法"、"中图法—主题词"、"科图法—主题词"、"主题词—中图法"、"主题词—科图法"等6个按钮选项。可按需要单击任一按钮完成相应转换,转换结束可对结果进行保存、浏览和编辑,并可为结果补充对应类号(第二分类号)。

5 系统功能评价

5.1 系统所完成功能

(1)从 MARC 文件或外部库文件转换为系统所需库文件。

(2)完成《中图法》、《科图法》与《汉表》之间共计等6项双向对照索引的生成。

(3)可人工选择标引词和分类号,便于对生成结果进行控制。

(4)保留了外部接口,用户可以方便地补充分类号对应的类名等事项。

5.2 系统评价

该系统较圆满地完成了用户所提出的需求,从程序运行的结果及用户对结果的分析统计来看,机编版与手工版相比较,二者相符率较高,二者不相符的部分,其中一部分系机读数据中错误的分类标引和主题标引数据,可以通过统计方法筛除;其他部分则可以弥补手工版中的遗漏和不足,克服手工版的冗余及错误。该系统可以排除标引人员手工对应标引的主观随意性,提高标引一致性、准确性及工作效率。

机编版的良好兼容性和使用的方便性解决了目前联机检索中所面临的不同的检索语言之间的兼互与互换方面的困难。

从软件角度来讲,系统采用面向对象的编程语言和事件驱动程序方式,界面美观友好,操作方便简单,维护和升级也较方便。

5.3 存在问题

该系统存在着下列一些主要问题:

(1)在处理双类号(即一条书目数据有两个中图法号或科图法号)对应多个主题词串时,将其交叉搭配生成多个记录。需要手工分析剔除其中的错误的记录。

(2)目前无法接收 Informix,Oracle,SQL Server 等大型数据库的数据。

(3)系统中许多程序算法略显复杂,系统的健壮性也不强。

6 结语

这套软件的开发，初步完成了分类号与主题词之间的双向对应标引，但下一步需研究如何利用统计学方法和语料库语言学的研究成果，通过对标引频次的统计分析，自动地剔除错误的数据，减少人工判别筛选的智力劳动。若能将其与自然语言(关键词)结合起来，建立关键词与对应主题词之间的转换，即自然语言与主题语言、分类语言之间的转换，这套系统将会更加完善，在建立中文情报检索语言转换系统方面发挥更大的作用。

参考文献：

[1]侯汉清.《中国分类主题词表》的原理、应用及改进.江苏高等学校图书馆学报，1998(1)

[2]傅守灿.图书馆自动化基础教程.北京：北京大学出版社，1996

[3]史济民.软件工程原理、方法与应用.北京：高等教育出版社，1992

[4]侯汉清.当代分类法主题法索引法研究.北京：书目文献出版社，1993

[5]安树兰等.中国机读目录格式使用手册.北京：华艺出版社，1995

分类号—主题词对照数据库的设计

南京农业大学信息管理系 张雪英 侯汉清

[摘要]本文通过调查分析我国机读文献数据库的标引数据,论述了我国尝试自动编制类表和词表转换系统的必要性和可行性。本文以我国现有的部分机读文献数据库的标引数据为实验样本,应用最大似然法,通过程序自动设计分类号—主题词对照数据库,以实现两种情报检索语言的兼容互换。

[关键词]最大似然法 情报检索语言兼容性 分类法 叙词表

我国对检索语言兼容性研究起步较晚。80年代后期,我国学者才开始探讨分类法与主题法的兼容与互换,即分类主题词表的制作。最有代表性的是《中国分类主题词表》(简称中分表)的问世。该表是在《中国图书馆图书分类法》与《汉语主题词表》的基础上编制而成的一部分类主题一体化词表,由160多位专家,40多家单位,历经8年编制而成。自1994年该表问世以来,已经历了4年的使用与验证。但是,随着文献的增加,新事物、新名词的不断涌现,其在实际应用中也暴露出许多缺陷:类目设置固定,更新周期长;有的类目对应词汇过多,出现冗余;有的类目对应词汇过于贫乏。主要原因是词表编制时以工作人员的主观性分析为基础,进行人工对应,文献保障性较差。因此,《中分表》还有待进一步改进。探讨高效率的、高性能的编表方法势在必行,以适应新的标引和检索需求。

本文拟采用统计方法,利用机读文献数据库资源,通过程序自动生成分类号—主题词对照数据库。

本文仅以中图法F83大类(金融,银行)为实验对象,所有程序用Visual FOXPRO 5.0实现。

1 实验数据的选取和分析

1.1 实验选材背景

我国中文机读文献数据库(主要调查含社会科学文献的大型数据库)及其标引形式如下:

表1 我国中文机读文献数据库及其标引形式

数据库	文献类型	标引形式
中文社科报刊篇名数据库	报刊	分类标引,主题标引(无主题词串)
中国国家书目(MARC)	图书	分类标引,主题词标引(含主题词串)
经济学科论文数据库	报刊,图书	关键词标引
中文科技期刊数据库	报刊	关键词标引
部分图书馆制作的MARC	图书	分类标引,主题词标引(含主题词串)

(1)主题标引数据分析：上述文献数据库的主题标引形式包括两种：含主题词串标引和不含主题词串标引(单个主题词标引)。一般说来，单个主题词很难准确地、完整地表达一个文献主题，在大多数情况下，只有采用主题词串形式才能满足标引和检索的需求。目前，我国只有中国国家书目(MARC)和部分图书馆制作的MARC数据库同时进行分类标引和含主题词串的主题标引。从理论上讲，这些标引数据可以直接用于主题标引和分类标引的对应转换。但事实上难以奏效，主要是因为这些数据含有大量的无效数据，从而难以有效地指导主题标引和分类标引的对应转换。本文拟采用统计方法，自动去除其中的绝大部分无效数据，重新生成一个有效的分类号－主题词对照数据库。

(2)标引对象分析：MARC数据的标引对象主要是图书。因此，根据MARC标引数据生成的分类号—主题词对照数据库不能适应报刊文献的标引。因为报刊文献的主题比图书文献要细化得多。一个能够有效地适应机器标引和检索的分类号—主题词对照数据库，必须同时满足图书文献和报刊文献的标引和检索。而我国的报刊文献数据库往往只有分类标引、关键词标引或不含主题词串的主题标引。因此，根据这些数据库直接生成的分类号—主题词对照数据库难以同报刊文献的分类标引和主题标引数据进行对应转换。本文试以《中文社科报刊篇名数据库》和《中文科技期刊数据库》两种较有代表性的大型报刊文献数据库为数据源，探讨如何自动生成分类号—主题词(串)对照数据库。

·《中文社科报刊篇名数据库》(简称上海库)由上海图书馆文达信息公司《全国报刊索引》编辑部负责研制和编辑。1993－1996年数据库共收录全国哲学社会科学期刊4500种，报纸170余种，基本上覆盖了全国邮发和非邮发的报刊。内容涉及马列主义、毛泽东思想、哲学、社会科学、及人文科学等各个学科，数据量超过72万条。使用《中图法》进行分类标引，参照《中国分类主题词表》进行主题标引。

·《中文科技期刊数据库》(简称中刊库)由国家科委西南信息中心编制。1989年开始创建，收录期刊约6000余种，数据量达200多万条，其中经济类数据达20多万条。中刊库已成为我国一个大型的综合性文献检索系统。使用《中图法》进行分类标引，采用关键词进行主题标引。

1.2 实验数据的获取

利用程序从两个数据库中下载F83类的所有记录(共52323条记录，其中上海库37654条，中刊库14669条)，抽取出分类号、主题词或者关键词字段(中刊库采用关键词标引，故将所有主题词和关键词统称为标引词)，生成数据库SJY(见附表一)。其记录格式如下：

分类号	标引词
F83	金融；竞争；研究
F83	银行；公共关系学
F83	金融；人才；培养；教学改革

＊注：中刊库标引词间隔符号为空格

进行主题标引和关键词标引时，通常按一定的次序来组织标引词，形成标引款目。但一个分类号往往对应着多个标引词，只有一个标引词或几个标引词的组合与分类号才是相对应的。

例：　标引款目　F83　金融；人才；培养；教学改革

标引款目中，只有“金融”、“人才”和“培养”三个标引词组合与F83类是相直接对应的。基于上述假设，本文通过程序，按照一定的模式将每条记录的分类号与该条记录的标引词进行匹配，生成四个样本库：

样本库1(YBK1):分类号与每个标引词一一匹配。共有181023条记录。
样本库2(YBK2):分类号与所有两个相邻标引词进行匹配。共有129369条记录。
样本库3(YBK3):分类号与所有三个相邻标引词进行匹配。共有77699条记录。
样本库4(YBK4):分类号与所有四个相邻标引词进行匹配,共有35084条记录。

YBK1记录格式:

分类号	标引词
F83	金融
F83	竞争
F83	风险分析
F83	银行
F83	公共关系学
F83	金融
F83	人才
F83	培养
F83	教学改革

YBK2记录格式:

分类号	标引词1	标引词2
F83	金融	竞争
F83	竞争	风险分析
F83	银行	公共关系学
F83	金融	人才
F83	人才	培养
F83	培养	教学改革

YBK3记录格式:

分类号	标引词1	标引词2	标引词3
F83	金融	竞争	风险分析
F83	金融	人才	培养
F83	人才	培养	教学改革

YBK4记录格式:

分类号	标引词1	标引词2	标引词3	标引词4
F83	金融	人才	培养	教学改革

2 信息对数量度和最大似然法的应用

从样本库的生成模式可以看出,每个样本库中包含大量的冗余信息,一个标引词(串)往往对应着多个分类号,一个分类号同样对应着多个标引词(串),本文的实验目的是要为每个标引词(串)确定一个正确的分类号。从统计学的角度看,每个标引词(串)与其对应的各个分类号具有不同的相关程度,也就是具有不同出现概率,相关程度越大,表明二者共现的概率越大,因此,选择相关程度最大的记录所对应的分类号为该标引词(串)对应的正确分类号。

统计学上要进行两个成对事件之间的相关度测量,必须首先统计它们的出现频次,其可能出现频次如下表所示:

表 2　两个成对事件的出现频次

	Yj	Yj'(j'≠j)
Xi	a	b
Xi'(i'≠i)	c	d

其中：

X 代表事件 A

Y 代表事件 B

a 代表 x_i 和 y_j 共同出现的频率

b 代表 x_i出现而 y_j 不出现(即 $y_{j'}$)的出现频率

c 代表 x_i不出现(即 $x_{i'}$)而 y_j 出现的出现频率

d 代表 x_i不出现(即 $x_{i'}$)和 Y_j 不出现(即 $y_{j'}$)的出现频率

a + b = Y_j 在样本中的出现频次

c + d = X_i在样本中的出现频次

统计学上有很多指标可用于测量两个成对事件的相关程度，较为广泛使用的主要有下面两种：

$IM(x_i;y_j) = aloga/(a+b)(a+c)$　　　(公式 1)

$Loglike = aloga + blogb + clogc + dlogd - (a+b)log(a+b) - (a+c)log(a+c) - (b+d)log(b+d) - (c+d)log(c+d) + (a+b+c+d)log(a+b+c+d)$　　　(公式 2)

例：假设在 YBK1 中，设 X 代表事件分类号，Y 代表事件标引词，标引词"金融"的出现频次为 80，分类号"F83"的出现频次为 300，分类号"F83"和标引词"金融"同时出现的频率为 30，则标引词"金融"和分类号"F83"的互信息量 IM = 30 × Log(30/80 × 300) = -6.68 × 30 = -200.4。

下面以 YBK1 为例，介绍具体的操作步骤：

(1)将样本库自动去掉重复记录(结果见表 3)。

表 3　样本库分类号和标引词情况一览表

数据库	样本总记录数	去重样本记录数	分类号个数	标引词个数
YBK1	181023	32529	366	13775
YBK2	129369	92677	366	52833
YBK3	77699	61736	351	58163
YBK4	35233	32085	295	31677

(2)根据公式 1 的要求，由程序生成每条记录的 IM 值(结果见表 4)。

表 4　样本库 1 数据记录

分类号	标引词	共现频次	分类号频次	标引词频次	Logl	IM
FL	ZTC	GXPC(a)	FLHPC(a+c)	ZTPC(a+b)	LogL	IM
F831	国际资本	2	2332	20	2.4362	-20.1141
F831.5	国际资本	1	1848	20	0.8093	-10.5176
F831.6	国际资本	11	1958	20	* 36.1589	* -89.9527
F831.7	国际资本	4	279	20	15.9457	-28.9626
F837	国际资本	1	560	20	1.8678	-9.3237
F837.125	国际资本	1	4	20	6.8870	-4.3820

分类号	标引词	共现频次	分类号频次	标引词频次	Logl	IM
F831	国际资金	3	2332	11	6.7157	27.1613
F831.5	国际资金	1	1848	11	1.3363	-9 9198
F831.7	国际资金	6	279	11	*31.3443	* -37.4241
F832.3	国际资金	1	1773	11	1.3736	-9.8783

(3)根据其IM值的大小,确定各分类号与该标引词的关联程度,选取IM值绝对值最大者为保留记录,即为每一个标引词(非重复标引词)确定了一个最佳分类号。如标引词"国际资本"的分类号确定为"F831.6",标引词"国际资金"的分类号为"F831.7"。

(4)由人工从YBK1中抽取出含有与F83类有关的标引词的记录,并选择了162条与F83类直接相关的记录进行统计分析,分类号与标引词对应正确率达80.3%。

(5)从第3步起,改用公式2,计算每条记录的Logl值(结果见表4)。用程序自动地为每个标引词确定一个最佳分类号,即logl值最大的记录。如"国际资本"的分类号确定为F831.6,"国际资金"的分类号为F831.7。本文共抽取203个与F83类有关的标引词进行抽样统计分析,其中分类号与标引词对应正确率达85.5%。

从上面的实验可以看出,logl方法的实验效果较为理想。主要原因有两个:一是样本量较小,Logl方法比较适合小样本的统计分析;二是分类号和主题词的共现频率较低,50%左右的记录共现频次小于3,采用logl方法,可以避免高Logl值携带较小的信息噪声。因此,本文拟采用logl方法,为每个样本库中的各个不同的标引词(串)均确定一个分类号,分类号与标引词(串)的相符率见表5。

表5 YBK库分类号与标引词(串)相符率

数据库	分类号与标引词相符率
YBK1	85.5%
YBK2	85.7%
YBK3	87.4%
YBK4	88.1%

在实验过程中,我们发现,分类号与标引词相符率随标引词串中标引词个数的增加而逐步递增,主要原因:一是标引词串中标引词个数越多,越能准确地表达文献主题,与分类号的相关程度较高;二是标引词串所包含的标引词越多,其出现频率越低,如YBK4中,大部分记录共现频率小于3。

3 正式分类号——标引词库的生成

经过上述处理,每个样本库中的每个标引词(串)都确定了一个分类号,下一步工作就是将每个样本库中与F83类有关的标引词(串)选择出来,这一步工作必须依靠人工辅助来实现。

1.制定筛选标准:

· 标引词(串)出现频次至少大于1

· 与F83类有关

· 分类号与标引词(串)对应正确

应用该标准可筛选掉大部分通用词、外类词和错误记录。

2. 从 YBK1、YBK2、YBK3 和 YBK4 四个数据库中手工筛选出 2282 条与 F83 类各类目相符的记录，组成新的数据库（简称 BZK－1）。

3. 将 BZK-1 按标引词排序，删除冗余记录和错误记录，包括单个标引词与标引词串之间的优选和多个标引词串之间的优选。例：

(1) F832.7　金融事业　　　　(删除)
　　F832.7　金融事业　中国　(保留)

(2) F831.2　金融体制　经济体制改革　　　(删除)
　　F831.2　金融体制　经济体制改革　中国　　(保留)

(3) F832 银行　关系　企业　　　　　　(删除)
　　F832 银行　关系　企业　中国　　(保留)

(4) F830.9　投机交易　　　　　(删除)
　　F830.9　投机　　交易　　　(保留)

第(4)种情况的出现主要是因为数据源有的采用主题受控标引，有的采用关键词标引，所以粗略选词时将这类标引词和标引词串视为等价标引词，然后再参照它们各自的出现频次，并结合人工判断进行筛选。这部分工作量较小，对于较有经验的标引人员来说，极易作出选择。

本文共选择出 1662 条记录(包括标引词串)，筛选率约为 1%(因 YBK 库中包含大量的非专业词汇，并且大多数非专业词汇出现频次较低，按筛选标准即被自动筛选掉，所以人工再选择花费的精力较少)。标准库 1(BZK－1)记录格式如下(记录片段见附表 2)：

分类号	标引词 1	标引词 2	标引词 3	标引词 4
F832.22	存款	中国		
F832.4	贷款	担保	中国	
F832.38	当铺			
F832.38	典当业	中国		
F831.5	股票市场	世界		
F832.49	国际租赁	中国		
F830.5	国家信用			
F832	银行	关系	企业	中国

4. BZK-1 中标引词的规范化。

由于标准库 1 中的部分数据来源是关键词标引，在标准库的生成过程中虽然进行过部分筛选，但为了确保数据的准确性和规范化，必须对标准库中的部分标引词规范为正式叙词。

(1)从《汉语主题词表》中抽取出与 F83 类有关的所有叙词和非叙词，生成汉语主题词表数据库(简称 CT)。为方便查找与 F83 类有关叙词和非叙词，本文先根据《中国分类主题词表》第 1 卷中各个分类号所对应的主题词，在第 2 卷找到相应的款目，然后建立成数据库格式，数据库结构如下：

字段	数据
叙词及非叙词	金融公司
分类号	F830.3
用项	
代项	财务公司
族项	公司
分项	

参项　　　　　　　投资公司,投资银行

(2)将CT库中有用项和代项的叙词(非叙词)与标准库进行自动匹配,自动将非叙词转换成正式叙词。标引词转换为正式主题词。部分未转换的关键词则自动上升为主题词。

(3)将BZK-1中一条记录对应的多个主题词组合成主题词串格式。

BZK-1转换为正式的分类号——主题词对照数据库,记录格式如下(参见附表2):

分类号	标引词
F832.22	存款-中国
F832.4	贷款-担保-中国
F832.38	当铺
F832.38	典当业-中国
F831.5	股票市场-世界
F832.49	国际租赁-中国
F830.5	国家信用
F832	银行-关系-企业-中国

4 从MARC记录生成分类号—主题词校正库

MARC库中标引数据存在较多的冗余信息和错误信息,必须将其优化,才可用于有效地指导分类与主题标引的自动转换。

本文仍采用LogL方法进行MARC标引数据的筛选,具体操作步骤如下:

(1)从MARC记录中抽取出分类号和主题词字段,生成分类号与主题词一一对应库(简称MARC库),本文只选择第一个主题词(串)。一般说来,第一个主题词(串)最能确切地表达文献主题。

实验数据:国家新书目(MARC)

样本空间:4138条记录

主题范围:F83金融;银行

(2)将MARC库去掉完全重复的记录(1482条无重复记录),然后分别求每条记录中分类号和主题词(串)的出现频次。

(3)应用公式2求出每条记录的LogL值,根据LogL值的大小为每个主题词(串)确定一个分类号。最后生成一个分类号-主题词对照数据库(简称BZK-2),筛选率为45%。

(4)对BZK-2进行人工筛选。记录片段参见附表3。

将BZK-1和BZK-2合并,生成一个正式的分类号与主题词对照数据库(简称DZK库)。同时,为弥补以上两个数据库的不足,将《中分表》补充至DZK库中。再次整理DZK库,共生成2463条记录。记录片断如下:

分类号	主题词
F832.22	储蓄-利率-中国
F832.22	储蓄-利息-中国
F830.48	储蓄-社会学
F832.22	储蓄-通货膨胀-中国
F832.22	储蓄-投资-中国
F830.48	储蓄系统

F832.22	储蓄－消费 中国
F832.22	储蓄－银行－中国
F832.22	储蓄－增长－中国
F832.22	储蓄－中国
F830.42	存贷业务核算
F832.23	存款－保险－中国
F830.48	存款乘数公式
F832.22	存款－分析－中国

5 结束语

随着计算机的广泛普及和文献数据库的急剧增长，标引人员和检索人员对情报检索语言提出了新的需求，一方面标引人员需要提高标引效率和标引质量，另一方面检索人员不再是专业人员，而是普遍的计算机用户，他们只希望用最简单的方法便能达到满足的检索效果。显然手工编制词表和词表对照表不仅费时费力，而且文献保障性差，应用效果也很不理想。探讨新型的词表兼容方法迫在眉睫。本文只是一种尝试性的试验，试验范围较小，所用方法还亟待进一步的研究和论证。

附表1:SJY数据库片段

F830.46 国际贸易 银行 信用证 结算 UCP500
F830.46 商业信用 银行信用 结算方式 国际贸易
F830.46 备用信用证 涉外经济活动 国际贸易 信用证 银行
F830.46 信用证软条款 识别 防范 信用证
F830.46 信用证 诈骗种类 防范 对策 银行 结算
F830.46 审单 付款 UCP500 信用证 银行 国际贸易
F830.46 信用证 UCP500 条款 银行
F830.46 信用证 可转信用证 对背信用证 对开信用证 银行
F830.46 银行 资金 结算 效率
F830.46 商业承兑汇票 银行 结算 企业 推广 对策
F830.46 信用卡 结算 中国 外国 银行
F830.46 银行 结算 改革 市场经济
F830.46 金穗卡业务 信用卡 银行 创新 结算
F830.46 信用卡 市场营销 风险控制 银行 结算
F830.46 信用卡 发行 使用 银行 美国
F830.46 金穗卡市场 信用卡 银行 结算 中国
F830.46 金穗卡 信用卡 非实时通讯网络 设计 银行
F830.46 商业票据化 金融 宏观调控 商业结算
F830.46 信用证 UCP500 银行 结算
F830.46 跟单信用证 信用证 银行 对外贸易 UCP500
F830.46 银行 结算
F830.46 银行 结算 资金周转
F830.46 信用证 结算 提单
F830.46 银行 结算
F830.46 商业汇票 结算 银行
F830.46 商业汇票 银行 结算
F830.46 银行 结算管理 企业
F830.46 银行 结算 资金
F830.46 企业 经营机制 银行 结算
F830.46 信用卡 银行 结算 中国
F830.46 银行 结算 中国 会计
F830.46 银行 结算 金融
F830.46 信用卡 银行
F830.46 银行 信用卡 专业银行
F830.46 银行 结算
F830.46 农业税 结算方式 银行 粮食企业
F830.46 托收承付 结算 银行 中国
F830.46 银行 结算 中国
F830.46 建设银行 信用卡
F830.46 银行 企业 结算
F830.46 银行 结算 柜员制
F830.46 银行 结算方式 改革
F830.46 信用卡 结算 银行
F830.46 建设银行 微机 建筑工程 结算
F830.46 银行 资金 结算
F830.46 信用卡 建设银行 结算
F830.46 信用卡 市场 推广 效果评价
F830.46 信用卡 结算 银行 中国
F830.46 信用卡 市场 建设银行
F830.46 期货交易 结算
F830.46 托收承付结算 结算 银行
F830.46 银行 结算 市场经济
F830.46 结算 工商银行 银行 监督 考核
F830.46 结算 银行 中国
F830.46 信用证 结算 债务 企业 银行
F830.46 监督 银行 结算
F830.46 承兑汇票 结算 银行
F830.46 结算 托收承付 银行
F830.46 信用证 时间概念 银行
F830.46 出口贸易 信用证 结算 单据 银行
F830.46 信用证 结算 风险 银行
F830.46 提单抬头 委托收款方式 银行 结算
F830.46 信用证 结算方式 单据 货物
F830.46 信用证 条款 银行
F830.46 出口企业 信用证 审证
F830.46 信用证 结算 风险防范
F830.46 跟单信用证 银行 处理规则 信用证
F830.46 信用证 规则 单据
F830.46 跟单信用证 信用证 规则
F830.46 工商银行 信用卡
F830.46 银行 汇票 结算 汇差
F830.46 联营机制 信用卡
F830.46 票据 银行 结算
F830.46 信用卡 中国银行 银行 北京
F830.46 信用卡 结算 透支风险 银行
F830.46 信用卡 人民币 结算 银行
F830.46 信用卡 透支 银行
F830.46 结算 对外贸易 中国 东欧 独联体
F830.46 债务 结算 银行
F830.46 信用证 UCP500 单证实务
F830.46 信用证 结算 银行 国际惯例
F830.46 国际结算 银行业务
F830.46 银行 结算 企业 债务
F830.46 商业汇票 对策 银行 结算

F830.46 信用卡业务 银行 竞争
F830.46 结算 银行
F830.46 银行 汇票 结算 资金 清算方式
F830.46 银行 结算
F830.46 银行 结算
F830.46 结算 银行
F830.46 银行 结算
F830.46 银行 结算
F830.46 工商银行 支票 结算
F830.46 结算 银行
F830.46 结算 风险 银行
F830.46 对外贸易 结算 银行 新疆
F830.46 出口商品 托收 结算方式 收汇
F830.46 结算备付金 银行 企业
F830.46 结算 托收承付 金融
F830.46 银行 结算 信用
F830.46 银行 结算 资金
F830.46 结算 资金 银行
F830 46 结算 托收承付 负效应 银行
F830.46 银行 结算
F830.46 银行 结算
F830.46 银行 结算
F830.46 银行 结算 经济效益
F830.46 结算 银行 对策
F830.46 结算 银行 对策
F830.46 信贷 贷款 进口贸易 结算
F830.46 银行汇票 企业 结算
F830.46 结算 银行 托收承付
F830.46 银行 结算
F830.46 银行 企业 结算管理
F830.46 信用卡 人民币 结算 银行
F830.46 进口业务 结算 货币 金融
F830.46 银行 帐务处理 分配 结算
F830.46 银行 结算管理
F830.46 结算管理 银行
F830.46 结算管理 银行
F830.46 结算管理 银行
F830 46 结算管理 票据 银行
F830 46 结算 银行
F830.46 结算 银行
F830.46 结算 贷款 银行
F830.46 银行 结算制度
F830.46 商业汇票 结算 会计处理 银行
F830.46 银行业务 结算
F830.46 银行 结算
F830.46 托收承付结算 结算 银行
F830.46 商业 汇票 承兑 结算
F830.46 支票 转帐 银行 结算
F830.46 结算 银行
F830.46 结算 结算式 银行业务
F830.46 结算 方式 银行业务 企业 资金核算
F830.46 票据 核算 结算 期票 汇票
F830.46 银行 结算 中国
F830.46 银行 结算 票据实收 贴现款额 计算
F830.46 银行 结算 期票 汇票 中国
F830.46 银行 结算办法 体制改革
F830.46 银行 清理 三角债
F830.46 银行业务 结算制度 中国
F830.46 银行 结算制度
F830.46 银行 结算 改革 安全
F830.46 结算 改革 银行
F830.46 银行 结算 改革 中国
F830.46 签名验证系统 数字银行 在线
F830.46 银行 结算 联行结算 建设银行
F830.46 结算资金 商品 流通 费用
F832.5 远期外汇交易 会计 理论 金融
F832.5 货币市场 金融 宏观调控机制 中国 中央银行
F832.5 农业银行 公开市场业务 金融市场 银行
F832.5 人民银行 公开市场业务 金融市场 银行
F832.5 保值贴补率 国债市场 中国 金融市场 储蓄
F832.5 离岸金融市场 银行 离岸金融业务 金融市场 中国
F832.5 中国 证券业 银行业 分业管理模式
F832.5 国有商业银行 金融市场 商品市场 竞争力
F832.5 同业拆借市场 商业银行 管理 效应 金融市场
F832.5 证券业 发展 建设银行 战略措施 证券市场
F832.5 中央银行 公开市场业务 中国
F832.5 共同基金 金融市场 中国

附表 2:标准库 1(BZK-1)记录片段

F830.46 货币结算
F830.46 货币结算-银行

F830.46 票据－银行
F830.46 期票
F830.46 人民币－货币结算
F830.46 商业承兑汇票
F830.46 透支
F830.46 托收承付
F830.46 信用卡－货币结算
F830.46 信用卡－透支
F830.46 信用卡－银行
F830.46 信用证－银行
F830.46 银行结算梗阻
F830.46 支票
F830.46 资金－货币结算
F830.46 银行－货币结算－改革－中国
F830.46 对背信用证
F830.46 对开信用证
F830.46 货币结算
F830.46 可转信用证
F830.46 商业票据化
F830.46 数字银行
F830.46 资金核算
F830.46 转帐
F830.46 货币结算
F830.46 商业结算
F830.46 收汇
F830.46 透支风险
F830.46 货币结算
F830.46 转帐结算
F830.46 货币结算
F830.46 货币结算
F832.5 拆借市场－中国
F832.5 公开市场业务－宏观调控－中国
F832.5 公开市场业务－中国
F832.5 公债－证券市场－中国
F832.5 股票发行－中国
F832.5 股票价格－中国
F832.5 股票交易－中国
F832.5 股票－金融市场－中国
F832.5 股票－流通－中国
F832.5 股票热－中国
F832.5 股票－上海市
F832.5 股票市场－发展－中国
F832.5 股票市场－上海市
F832.5 股票市场－深圳市
F832.5 股票市场－台湾
F832.5 股票－证券交易－中国
F832.5 股票－中国
F832.5 股市－中国
F832.5 国库券－金融市场－中国
F832.5 国有股－流通－中国
F832.5 国有资产－中国
F832.5 国债市场－中国
F832.5 黄金市场－中国
F832.5 汇率－香港
F832.5 金融市场－发展－中国
F832.5 金融市场－公开市场业务－中国
F832.5 金融市场－股票市场－中国
F832.5 金融市场－江西省
F832.5 金融市场－期货交易
F832.5 金融市场－上海市
F832.5 金融市场－社会主义
F832.5 金融市场－深圳市
F832.5 金融市场－证券交易－中国
F832.5 金融市场－证券市场－中国
F832.5 金融市场－中国
F832.5 金融衍生工具－中国
F832.5 竞价发行－中国
F832.5 可转换债券－中国
F832.5 离岸金融市场－中国
F832.5 票据－市场－中国
F832.5 期货交易－中国
F832.5 期货市场－风险管理－中国
F832.5 期货市场－中国
F832.5 上海证券交易所
F832.5 上市公司－中国
F832.5 同业拆借市场－中国
F832.5 外汇市场－发展－中国
F832.5 外汇市场－中国
F832.5 银行同业拆放市场－中国
F832.5 有价证券－中国
F832.5 债券市场－中国
F832.5 债券－中国
F832.5 证券回购市场－中国
F832.5 证券交易－市场－中国
F832.5 证券交易所－中国
F832.5 证券交易－中国
F832.5 证券评级制度－中国
F832.5 证券市场－福建省

F832.5　证券市场－管理－中国
F832.5　证券市场－规范化－中国
F832.5　证券市场－山东省
F832.5　证券市场－投资风险－中国
F832.5　证券市场－中国
F832.5　中国－B股市场
F832.5　资本市场－发展－中国
F832.5　资本市场－融资－中国
F832.5　资本市场－香港
F832.5　资本市场－中国
F832.5　资本市场－规范－发展－中国
F832.5　股票－分析－研究－中国
F832.5　银行－储蓄－存款－中国
F832.5　证券市场－监督－管理－中国
F832.5　证券市场－问题－对策－中国
F832.5　证券市场－发展－问题－中国
F832.5　证券市场－中国－人物－访谈录
F832.5　证券市场－国际化－策略－中国

附表3:标准库2(BZK－2)记录片段:

F830.46　非现金结算
F830.46　结算业务核算
F830.46　票据
F830.46　票据－结算业务核算
F830.46　企业－结算业务核算
F830.46　商业－汇票－承兑
F830.46　同业往来
F830.46　信用卡
F830.46　信用证－贸易
F830.46　智能卡
F832.5　公债
F832.5　公债－证券交易
F832.5　公债－资本市场
F832.5　股份
F832.5　股份有限公司
F832.5　股票－交易
F832.5　股票－证券交易
F832.5　农产品－期货交易
F832.5　期货交易－企业管理
F832.5　企业－房地产
F832.5　企业－信贷资金
F832.5　社会主义经济－资本市场
F832.5　证券
F832.5　证券交易所
F832.5　资本市场
F832.5－66　证券交易所－统计资料
F832.5－66　资本市场－统计资料

参考文献:

[1]Lois Mai Chan: Feasibility of a computer generated subject validation file base on frequency of occurrence of assigned LC subject headings.
Http://www.oclc.org/oclc/research/publications/review95/part2/chan

[2]Diane Vizine－Goetz: Subject headings for everyone: Popular library of congress subject headings with Dewey Numbers OCLC Newsletter May/June,1998

[3]Daille: Study and Implementation of combined techniques for automatic extraction of terminology, MIT press,1996

[4]李波.计算机建立分类法-叙词表转换系统的尝试.南京农业大学信息管理系毕业论文,1998

军用主题词表系列化建设现状与发展

全军军用主题词表编制管理委员会办公室　曹　杰　赵建国

[摘要]简要介绍了军用主题词表系列化建设的基本作法，已完成的19部专业表的基本情况。同时对系列化建设今后的发展方向和主要任务做了概括叙述。

[关键词]军用主题词表　系列化　基本情况　发表方向

军用主题词表系列化建设(简称军表系列化建设)是指在编制《军用主题词表》(简称〈军表〉)这一供全军统一使用的大型综合性主题词表的基础上，为满足一些保密性、专业技术性较强的单位需要，编制部分与《军表》相互兼容、互为补充的专业性词表，并做为《军表》的分表，与《军表》共同构成军队统一的主题词表系列。这是一项在国内外主题词表编制中具有创造性的工作，它标示着我军主题检索语言的发展步入了系列化、兼容化的轨道。

1　军表系列化建设的基本做法

在《军表》编制之前，军队系统只有《军用公文主题词表》、《情报专业主题词表》、《后勤科学技术主题词表》和《国防科学技术叙词表》4部专业性词表。1988年《军表》编制工作展开后，全军有十几个单位在完成《军表》编制工作的同时，先后提出编制供本系统使用的专业词表。如何处理好《军表》与各专业表之间的关系，使二者具有良好的兼容性，保证全军主题检索语言的统一，是军表系列化建设的关键问题。对此，通过深入调查和分析国内外主题检索语言编制的经验和教训，与有关专家进行研究探讨，我们在实践中主要采取了以下基本做法。

1.1　确立了系列化建设的指导思想

编制《军表》的目的，就是要为军队指挥自动化建设和文献信息管理提供统一规范的检索语言，为实现文献信息的联机检索与资源共享打好基础。从这一点看，全军使用一部词表最有利于达到这一目的。但从实际出发看，尽管《军表》在收词范围和范畴设置上已充分考虑各学科专业和业务部门的需要，但对一些保密性、专业技术性较强的单位来说，仍难以满足需要。因此，在《军表》基础上，根据实际需要编制一定数量的专业性词表是完全必要的。编管会办公室经与全军各专业表编制单位充分协商，大家确定了军表系列化建设的指导思想，即：遵循共同的标准和规定，统一规划《军表》与专业表编制管理工作，实现两者之间的相互兼容和互为补充，共同形成统一规范的军队主题检索语言体系。这一指导思想的形成和贯彻，保证了军表系列化建设工作能够健康有序发展。

1.2　确定了专业表编制的基本原则

专业表编制涉及的单位多，各专业表在学科内容层次、涉及专业范围和使用对象等方面也不尽相同。如果要求各专业表均按照《军表》的模式统一编制，不仅难以实现而且也没有必要。如果由

各单位自行其事，则难以保证与《军表》兼容。为此，在1990年至1991年多次召开的专业表编制工作会议上，经广泛听取各单位意见和专家建议，确立了专业表编制应贯彻“宏观控制、微观灵活”的基本原则。“宏观控制”是指：第一，对专业表的编制实行报批制度，凡《军表》能够满足本单位文献信息管理需要的，原则上不再编制专业表；第二，专业表的编制要以《军用主题词表编制规则》为技术规范标准，在选词原则，词形、词义规范，用代关系等方面应与《军表》保持一致；“微观灵活”是指专业表编制在组织形式，词表结构，属分关系、相关关系，主题词款目格式及显示方式等方面，可根据需要灵活处理。这一原则既体现了军表系列化建设的指导思想，也使各专业表能够体现自身特点，因此，在专业表的编制中基本得以落实。

1.3 具体措施和方法

为使军表系列化建设能够实现预期目的，我们在工作实践中，主要采取了以下措施和方法。

(1)加强计划指导。为使专业表编制工作有章可循，在综合各专业表编制单位和有关专家意见基础上，全军军表编管委员会于1991年9月制定和下发了《军用专业主题词表编制工作暂行规定》，明确了军用主题词表系列的概念，《军表》与各专业表的关系，专业表编制的范围、原则以及管理等内容。同时根据各单位的安排，制定了军表系列化工作计划，及时了解掌握各专业表的编制情况和进度，协助解决有关的问题。

(2)完善技术标准。国家军用标准——《军用主题词表编制规则》是《军表》与专业表编制工作共同依据的技术规范标准。为保证该标准能够与《军表》和专业表的编制相适应，及时吸取各方面的经验，反映主题法理论研究与实践的新成果，在十年的时间里，我们对该标准进行了两次较大的修订。第一次是1991年《军表》编制工作完成后，为吸取《军表》编制的经验，对该标准(试行稿)进行了修订。第二次是1997年，为配合《军表》的修订工作，同时更好的指导专业表编制和修订，我们再次组织有关人员对该标准进行了修订，并于1998年完成。这次修订充分吸取了几年来专业表编制工作的经验，对有关技术内容作了较大改动，增加了专业表编制要求等内容，使其更具有针对性和可操作性，为军表系列化建设提供了技术标准依据。

(3)搞好服务保障。为解决各单位在专业表编制中人员水平不一，设备力量不足等问题，我们组织有关专家和办公室人员，对各单位的编表工作给予及时的技术指导，包括培训讲课、提供技术咨询、协助解决难点问题等。同时还组织开发了军用专业主题词表微机编制管理系统，这套系统于1992年10月投入使用，至今已协助有关单位完成了军用公文、军事情报、军用气象水文、军队档案、第二炮兵、军队政治工作、军事测绘与军事地理和通信兵等8部专业表的编制或修订工作，不仅使各单位提高了编表效率，而且进一步促进了军表系列化建设的统一规范。

(4)认真审查把关。军表编管委员会办公室对各单位的专业表负有审查和与编表单位共同组织技术鉴定的责任。在审查把关中，坚持以《军用主题词表编制工作暂行规定》和《军用主题词表编制规则》为依据，着重解决影响与《军表》兼容的词形词义、用代关系不规范等问题。对意见不一致的地方，则本着实事求是的态度协商解决。在指出和纠正专业表中某些不当之处的同时，也注意发现和记录《军表》中存在的问题，为《军表》的修订提供依据。从而较好地解决了《军表》与专业表之间相互兼容统一的问题，保证了军表系列化建设的质量。

2 军表系列化建设的成果

军表系列化建设自1988年开始至1998年底，全军已有22个单位完成了19部专业表的编制

或修订工作，其中1988年以后编制的15部，占总数的78.9%；目前正在编制的2部，即海军和电子对抗专业主题词表，预计2000年完成；在已完成的19部专业表中，有4部分别获得国家和军队科技进步奖。

下面对已完成的19部专业表情况分别做一简要介绍。

2.1 《军用公文主题词表》

该表由总参办公厅秘书局和总参通信部共同组织编制。主要供军队各级机关的军用公文和全军指挥自动化网上文电传输主题词标引使用。于1987年首次编制完成并试用，1991年经修订后正式颁发使用，1997年开始组织第二次修订，至1998年底已完成，并做为国家军用标准报批。修订后的词表共收词5834个，其中正式主题词5166个，非正式主题词668个，入口率为12.9%。全表由主表(字顺表)、附表和范畴索引组成，并附有《军队机关公文主题词标引规则》(GJB2418.5—95)。附表包括部队机关名称表、军队人员称谓表。

2.2 《军队档案常用主题词表》

该表由总参办公厅档案局和空军政治学院信息管理系共同编制。主要供全军档案业务部门对军队现行文书档案及其它类型档案文献的主题标引与检索使用，于1994年12月正式出版。该表共收词10916个，其中正式主题词9421个，非正式主题词1495个，入口率为15.9%。全表由主表(字顺表)、附表、范畴索引、轮排索引等组成。附表包括军队编制序列表，军队人员称谓表，中国行政区划表，常用时间术语表。在该表基础上已开发有电子版词表系统。

2.3 《军用气象水文主题词表》

该表由总参作战部气象局组织编制。主要供全军气象业务部门、院校和研究单位对气象水文资料及有关文献数据库的主题标引与检索使用，于1996年底完成并出版。该表共收词8085个，其中正式主题词5148个，非正式主题词2407个，组代关系词530个，入口率为57.1%。全表由主表(字顺表)、范畴索引、英汉及汉英对照索引、族首词索引组成。为方便使用，还编制了《军用气象水文主题词释义词典》和《军用气象水文主题词表使用说明》。

2.4 《军事情报主题词表》

该表由总参二部组织编制。主要供军队情报部门对军事情报资料及情报数据库的主题标引与检索使用。于1988年6月编制完成，1991年5月进行了修订再版，1993年11月进行了第二次修订后做为国家军用使用标准颁发，标准代号GJBz—20190—93，1998年9月开始组织第三次修订。该表(1993年版)共收词5760个，其中正式主题词5122个，非正式主题词638个，入口率12.5%。全表由主表(字顺表)、范畴索引、英汉对照索引(汉英对照随主表)组成。

2.5 《情报专业主题词表》

该表由总参三部九局组织编制。主要供三部系统技侦情报资料管理和情报数据库的主题标引与检索使用。该表是在1985年编制的《情报专业主题词典》(试用本)基础上，根据军表编制规定进行修订后，于1988年编制完成。1998年下半年开始组织第二次修订，预计1999年完成。该表(1988年版)共收词4859条，其中正式主题词4440条，非正式主题词419条，入口率9.4%。全表由主表(字顺表)、范畴索引、英汉对照索引(汉英对照随主表)组成。该表开发有用于情报数据库的

电子版词表。

2.6 《技侦技术专业主题词表》

该表由总参三部八局组织编制。主要供军队技侦系统统一规范有关技侦技术专业的主题标引、检索用语和建立自动化信息、情报管理体系使用，于 1993 年 12 月出版。该表共收词 10715 个，其中正式主题词 9971 个，非正式主题词 744 个，入口率 7.5%。全表由主表(字顺表)、附表、范畴索引、英汉及汉英对照索引组成。附表包括卫星、航天器、通信卫星地球站地名、密码机、密码机公司、世界各国行政区名称、有关国际组织机构等名称表。

2.7 《通信兵主题词表》

该表由总参通信部组织编制。主要供全军通信、指挥自动化部门和通信院校、研究单位对文献资料及有关信息数据库的主题标引与检索使用。于 1994 年完成(未正式出版)，1997 年起组织进行了较大的调整修订，至 1998 年底完成并通过审查，已做为国家军用标准报批。该表共收词 14928 个，其中正式主题词 8942 个，非正式主题词 5986 个，入口率 67%。全表由主表(字顺表)、范畴索引、英汉及汉英对照索引组成。为方便使用，该表的释义词典正在编制中，预计 1999 年完成。拟在通信公共信息网上使用的电子版词表目前也正在开发中。

2.8 《防护工程主题词表》

该表由总参工程兵第四设计研究所编制。主要供军队工程兵业务部门和有关院校、研究单位对工程防护专业文献资料及相关数据库的主题标引与检索使用，于 1990 年 8 月编制完成，1991 年出版。该表共收词 5576 个，其中正式主题词 5290 个，非正式主题词 286 个，入口率 5.4%。全表由主表(字顺表)、词族索引、范畴索引、英汉对照索引和机构表(附表)组成。

2.9 《防化兵主题词表》

该表由原总参防化部(现总参兵种部防化局)组织编制。主要供军队防化系统图书情报部门对专业文献资料主题标引与检索使用，于 1990 年 4 月出版，1998 年起，由防化学院进行修订。该表共收词 3187 个，其中正式主题词 2925 个，非正式主题词 262 个，入口率 8.9%。全表由主表(字顺表)、范畴索引、英汉及汉英对照索引组成。

2.10 《军用机要专业主题词表》

该表由总参机要局组织编制。主要供全军机要部门对机要电报及有关专业文献资料主题标引与检索使用，于 1993 年 10 月出版。该表共收词 2428 个，其中正式主题词 2112 个，非正式主题词 316 个，入口率 15%。全表由主表(字顺表)、附表、范畴索引、英汉及汉英对照索引组成。附表包括型号表、部队番号表。为方便使用，还编有《军用机要专业主题词释义词典》。该表因其收词内容的特殊性，被定为机密等级，只供机要部门内部使用。

2.11 《军事测绘与军事地理主题词表》

该表由总参测绘局组织编制。主要供全军测绘业务部门和院校、研究单位用于相关专业文献资料和数据库的主题标引与检索，于 1998 年 7 月出版。该表共收词 5913 个，其中正式主题词 3896 个，非正式主题词 2017 个，入口率 51.8%。全表由主表(字顺表)、附表、范畴索引、英汉及汉

英对照索引组成。附表包括国际测绘组织和国家测绘机构表,国外主要海、空军基地表。

2.12 《军队政治工作主题词表》

该表由总政办公厅组织编制。主要供全军政治部门和院校对军队政治工作领域各专业的文献资料档案和信息数据库的主题标引与检索使用,于1998年7月出版。该表共收词21125个,其中正式主题词8995个,非正式主题词3130个,入口率34.8%。全表由主表(字顺表)、附表、范畴索引组成。附表包括机构名称表、中国地区名称表、世界各国和地区名称表以及节日表。

2.13 《军用后勤主题词表》

该表由总后勤部组织编制。主要供全军后勤系统及院校、科研单位的各类文献资料管理和建立后勤系统的信息管理系统使用,于1992年5月出版。该表共收词14872个,其中正式主题词13886个,非正式主题词986个,入口率7.1%。全表由主表(字顺表)、范畴索引、英汉对照索引组成。

2.14 《后勤科学技术主题词表》

该表由总后勤部后勤科学研究所编制。主要供总后系统办公自动化、后勤技术文献主题标引与检索使用,于1980年首次编制完成,1988年经修订后再版。该表共收词4996个,其中正式主题词4225个,非正式主题词771个,入口率18.2%。全表由主表(字顺表)、范畴索引、英汉对照索引组成。

2.15 《军用医学主题词表》

该表由总后勤部卫生部和军事医学科学院共同组织编制。主要供军队卫生行政机构、医疗、预防、教学及科研单位对医学科学各专业和卫生行政管理方面的各类文献图书资料及数据库的主题标引与检索使用,于1993年8月出版。该表共收词23387个,其中正式主题词20662个,非正式主题词2725个,入口率13.2%。全表由主表(字顺表)、范畴索引、英汉对照索引组成。

2.16 《国防科学技术叙词表》

该表由原国防科工委情报所(现总装备部情报所)组织编制。主要供军队和地方有关国防科技情报部门对国防科学技术文献主题标引及国防科技信息网络建设使用。该表于1985年编制完成,1991年进行了全面修订后再版,1995年10月补充出版了型号附表。该表主表(字顺表)共收词31816个,其中正式主题词28996个,非正式主题词2820个,入口率9.7%。型号附表共收词22673个,其中正式主题词14454个,非正式主题词8219个,入口率56.9%。全表由主表(字顺表)、型号附表(后补充编制)、范畴索引、英汉及汉英对照索引组成。基于该表开发的电子版系统已于1997年完成并投入使用。

2.17 《军用航空主题词表》

该表由空军司令部组织编制。主要供空军机关、航空兵部队、院校、研究单位等对有关文件档案、信息资料、学术论文、教材等的主题标引与检索,于1992年6月编制完成。该表共收词4542个,其中正式主题词4041个,非正式主题词501个,入口率12.4%。全表由主表(字顺表)、范畴索引、英汉及汉英对照索引组成。

2.18 《机场工程主题词表》

该表由空军工程设计研究局组织编制。主要供空军机场、营房工程研究、设计、建设部门和单位对相关文献资料的主题标引及信息网络建设使用，亦可供民用航空工程建设部门参照使用，于1997年11月完成。该表共收词8181个，其中正式主题词5902个，非正式主题词2279个，入口率38.6%。全表由主表(字顺表)、范畴索引、英汉及汉英对照索引组成。

2.19 《第二炮兵主题词表》

该表由第二炮兵司令部组织编制。主要供第二炮兵各级机关、院校和科研单位对公文、档案、专业信息资料及学术论文、图书教材等的主题标引与检索使用，于1997年9月出版。该表共收词9302个，其中正式主题词8404个，非正式主题词898个，入口率10.7%。全表由主表(字顺表)、范畴索引、英汉及汉英对照索引(1999年9月单独出版)组成。另外配套编制的《第二炮兵主题词释义词典》已完成，预计1999年9月出版。

3 军表系列化建设今后的发展方向和主要任务

军表编管会办公室组织研制开发的《军表》应用管理系统，对传统文本词表进行了重大改革和创新，为计算机和网络环境下新型检索语言的设计提供了新思路。1998年9月，由国内多位知名专家组成的鉴定委员会对该项成果给予了高度评价，一致认为："该成果达到了当前国内外技术领先水平，标志我国叙词语言的研究和应用进入了一个新的发展阶段。"但目前全军现已完成的专业词表，绝大多数还停留在只适于手工标引的文本式词表阶段。因此，今后军表系列化建设的发展方向是：以《军表》应用管理系统推动和指导全军各专业表的电子化，进一步提高全军文献检索语言建设的整体水平，并在此基础上建立全军主题词库。根据这一方向，今后几年军表系列化建设的主要任务是：

3.1 抓紧部分专业词表的修订更新工作

对词表进行动态管理、适时更新，是其生命力的基本保证，也是开发电子版词表的基础工作。目前，全军完成的19部专业表中，有11部已达到修订时限，其中3部的修订工作已展开，其余8部应抓紧安排。修订工作以补充新词和大幅度提高词表入口率为重点。

3.2 指导和协助各专业表实现电子化

全军各专业主题词表电子化工作，要根据各单位的情况，分期分批进行，不搞一刀切。1999年下半年将召开全军专业表电子化工作会议，具体研究和部署，并进行试点。力争3—5年内全部实现电子化。

3.3 建立全军主题词库

在军表系列化建设和《军表》及专业表电子化的基础上开发建立全军主题词库，是我们今后工作的一项重要任务。通过建立全军主题词库，保证和实现全军在多数据库联网环境中检索语言的兼容和统一；同时为军事领域主题词表的编修以及辞书、术语、自然语言处理等相关课题研究提供基础数据。主题词库的结构设计，拟以《军表》电子版为主库，以《军表》系列专业表电子版为子库；

主库与子库可分可合，既可独立使用又可联合或混合使用。全库在网络上形成一个既开放又兼容的统一整体。建立词库的技术关键问题是解决好主库与子库挂接并实现联合或混合显示。

以上对军表系列化建设的基本情况做了一个简要概述。对这项工作今后的发展，我们还将通过这次研讨会进一步论证完善，使军表系列化建设能够持续深入的发展。在军表系列化建设中，钱起霖、张琪玉、付兰生、鲍绵福、刘湘生等各位专家对此项工作给予了大力支持和帮助，特此表示衷心感谢。主题词是主题词表最基础的组成单位，主题词的有效性和规范程度如何直接决定词表的质量。我们的设想是通过建立一个大型的主题词库，对主题词进行统一规范处理和及时的动态管理，库中的词将做为词表编制或修订的"原材料"、"标准件"提供给户。这不仅可以从根本上解决各表的兼容问题，而且也能够为解决现代信息处理技术飞速发展，新的名词术语不断出现，词表的编制修订总处于滞后状况这一矛盾，提供一种新的有效解决方法。特别是在词表实现电子化后，词库将及时向各词表提供经规范化处理后的新词，而词表也可将在使用中所遇到的欲增补的词传递给词库，进行规范处理。双方将形成一种互动关系，从而建立一种全新概念上的主题检索语言体系。建立主题词库的基础是要具备有一定数量上的、经过统一规范处理、具有合理用代关系的主题词。目前，全军已完成的专业表和《军表》的收词总量已达到262595个(含各表间的重词)，其中正式主题词219338个，非正式主题词43257个。今年《军表》修订工作完成后，收词量将进一步增加，特别是对新的名词术语的补充和非正式词的增加比例将更大。可以说，目前建立全军主题词库的基础已经具备，虽然这里面还要做大量的对词的规范处理和增补主题词各要素的工作，但由于这是一项对军队指挥办公自动化建设和现代文献信息管理具有长远意义的工作，因此，将成为我们今后一项长期的重点工作。

《军用主题词表》在文献自动标引检索服务系统中的运用

大连陆军学院　杜文芝　冯小平

[摘要]本文论述了利用《军用主题词表》开发研制“军事文献主题词自动标引检索系统”中词典库建立的原则，实现主题词自动标引的过程及检索方法等。

[关键词]军用主题词表　文献标引　自动标引　检索系统

《军用主题词表》(以下简称《军表》)是我军第一部大型、综合性词表，对实现我军检索语言统一，情报资源共享和加速我军指挥、办公自动化的过程具有重要意义，为实现自动化情报检索，建设全军统一的情报检索系统奠定了基础。我院图书馆于1996年利用《军表》完成了对“军事主题词自动标引检索系统”的开发研制工作。经实际运行取得了较好的效果。下面就《军表》在系统中的运用情况做以概述。

1　利用《军表》建立机内词典库

《军表》收词丰富而全面，知识覆盖面较宽，共收词52500条，是一部大型的军事词表，涉及87个学科专业。就机关、部队和院校而言，仅本系统、本部门、本单位应用《军表》，所涉及的专业范围是有限的。所以，在开发研制主题词检索系统中，在建立主题词机内词典库时，有必要根据需要来选择专业应用范围，研究如何建立机内词典库的问题。

1.1　按实际应用专业范围建库

我们在编制主题词标引检索系统时，无论采取怎样的方法实现主题词的标引和检索都需要建立主题词库。如考虑本系统、本部门、本单位应用，就必须从《军表》中准确地选择专业范围和相关学科的范畴，不必要把《军表》中所有的词条都建成词典库。否则一是，占数据空间相对太大；二是，明显地影响运行速度。我们可以根据学科的变化和发展随时增加学科范畴和增加词典库的数量。我院图书馆所开发的“军事文献主题词自动标引检索系统”目的是在全军初级陆军指挥院校范围中应用。此类院校性质相同，课程设置、专业范围、重点学科是一致的，入藏的核心期刊也是大致相同的。在此前提下，我们按初级陆军指挥院校所设置的“军事理论”、“战役战术”、“军队管理”、“军队训练”、“政治工作”5个重点学科，按照《军表》的“范畴表”分别建立了5个词典库。

1.2　《军表》词典库结构字段的设置

《军表》的范畴表反映了学科属性，分成三级范畴，基本上是按照习惯的学科划分而形成类目，类目大部分都是知识门类。

例如：

03 战役战术(一级范畴)

03C 作战样式(二级范畴)

03P 装甲兵战术(二级范畴)

03Q 工程兵战术(二级范畴)

03R 通信兵战术(二级范畴)

03S 电子对抗兵战术(二级范畴)

03T 防化兵战术(二级范畴)

所以,我们所建立的5个库是按学科属性的一级范畴建库的,我们在设置数据字段时给予了二级范畴和三级范畴号码的字段位置,进行了类目的标识。通过词典库结构字段的设置情况就可以看出,我们可以实现从主题词检索,又可以从范畴类目途径实现文献检索,这样建库便于系统按文献类别分头进行标引,专指度高,标引速度快。

1.3 建立文献标引辅库

我们所研制的"军事文献主题词自动标引检索系统",除建立机内《军表》五个重点学科词典库和文献主库等数据库外,我们并建立了文献标引辅库。此库的数据来源于往主库中新录入的一批文献,当往主库中录入一批文献后,系统自动把这批文献送给此库,当系统自动标引主题词时,主题词往此库中标引,标引完成后,系统会自动将这批数据追加到文献主库中去。然后,系统自动删除此库数据,等待下一批新文献的标引工作。我们这样设计数据库的优点很明显,即限制一批新文献数据在标引辅库中操作,可防止标引主题词时破坏主库,使其他数据受损,又可避免一篇文献被重复标引。《军表》词典库只是对辅库中的一批文献标题和文摘进行标引,所以运行速度快。明显地减轻了工作人员的劳动强度,节约了时间。因为一批数据数量有限,便于标引员进行数据质量检查和纠错,以及未被自动标引上的数据进行人机对话进行标引等。

2 利用《军表》实现对文献主题词的自动标引

近十几年来,语言学界和人工智能领域的学者在汉语自动分词与自动标引上进行了大量的研究,找到了许多解决汉语分词与标引的方法,这些方法实现的手段也不尽相同。常用的"微机辅助文献标引"是标引员通过词表接口用鼠标或击键选择标引词加人待标引的文件中;另外,还有应用词典切分法和切分标记法等。

2.1 主题词自动标引的实现方法

近几年来,军事用语逐渐规范化,文献标题军事用语率较高,我们在编制文摘时也尽量使用比较规范化的军事用语,这为提高文献的自动标引率打下了良好的基础。我们采用的方法是用机内主题词对文献标题和文摘进行匹配的方法,就是采用布尔逻辑运算,利用机内主题词对文献标题和文摘进行字符串一一匹配而得到主题词。被匹配上的字符串,系统会自动添加到标引主题词字段中去。采用以上这种抽词方法,首先能保证抽出的标引词是机内主题词典所包含的词,即为正式主题词,其次能够保证其语法比较正确。我们只需判断是否适用于本篇文献的主题即可。用此种方法不必抽取自由词,即词典中没有的词,也可以减少人工判断。为用户提供了与标引人员大致相同的文本环境,无疑奠定了标引与检索一致性的基础。减少了检索的不确定性,提高了系统的效率。不用标引工作单,标引工作直接在微机上进行,减少了原二次文献加工流程中不必要的中间环节和人为差错。由于计算机产生的标引词是机内主题词典中的词,因而保证了标引员优先选取这些词

作为最终标引词，不会产生自由词。匹配的字符从多至少，为了遵循长词优先的原则，若一个长词已经命中，则不用再抽取比较短的词。这样匹配命中的索引词添加到文献库中的主题词字段中去，以待检索之用。

2.2 《军表》中词间的关系转换

《军表》与其它叙词表相同，主题词是通过词间的“Y(用)”、“D(代)”、“S(属)”、“F(分)”、“C(参)”语义网络来展示词间关系的。这些关系在“文献自动标引系统”中需要进行恰当处理才能保证标引的准确率。

“Y(用)”符号后面的词为正式主题词，在建库时全部按正式主题词输入，经系统标引匹配命中的就作为正式主题词添加到主题词字段中去。

“D(代)”符号后面的为非正式主题词，在系统运行中遇到有“D”字符的就用正式主题词替换后添加到主题词字段中去。

“S(属)”，此符号后面的词为主题词的上位词；“F(分)”，此符号后面的词为主题词的下位词；“C(参)”，此符号后面的词为参照词。根据以上词间的关系进行处理，系统进行逻辑运算，产生并选取准确的主题词，从而解决以上词间各种关系的转换，达到选取主题词正确，标引准确的目的。

2.3 未自动标引上主题词的文献的处理

我们对每篇文献设置了4个主题词标引字段，系统对文献进行主题词自动标引工作完成后，会有一部分文献主题词字段标引得不完全，或全部没有被标引。原因是文献标题和文摘没有被主题词典库的词匹配上。这时标引员需逐条对数据进行检查，通过人机对话进行标引，将此批文献进行完全标引。为防止错标、漏标和过渡标引，我们需将辅助标引库中的数据逐条进行质量检查，对未被标引上或存在问题而标引不准确的数据进行添加 或修改，以保证主题词标引的准确率。

3 《军表》在系统检索中的运用

在对系统进行文献检索的编程中，必须树立用户第一的观念，要考虑用户使用方便，界面清晰，方便检索。这样，就必须加强编程技巧。我们采用了移动光标和使用下拉式菜单方式，不需要用户输入汉字和翻阅《军表》，即可实现文献检索的目的。

3.1 《军表》范畴类目检索

《军表》范畴类目从知识门类反映了学科属性。我们所编制的系统可实现从范畴进行文献检索的功能。我们将《军表》中有关初级陆军指挥院校所涉及的“军事理论”、“军队管理”、“战役战术”、“军队训练”、“政治工作”五个重点学科，做为五个“一级范畴”。当系统进入范畴检索后，屏幕显示这五个范畴的名称供选择。确认后，屏幕可显示所有此范畴的文献。当你选择某“一级范畴”中的某“二级范畴”中的某一类目时，确认后，这个“二级范畴”类目的文献立即被检索出来。依此类推，选择某一“二级范畴”中某一“三级范畴”类目后，可检索出这个“三级范畴”类目中的所有文献。系统在完成范畴类目文献查询检索中，反响速度快，可用秒计算。并可随意打印所查询出的一批文献。

3.2 主题词检索

本系统为用户设置和提供了可从范畴类目途径完成主题文献检索。当用户选择进入某一“三

级范畴"后，即可进行主题检索。每一范畴的主题词都按汉语拼音的顺序排列，显示在屏幕上供用户查询。只需移动光标到要查找的主题词处，点击鼠标或按回车，立即显示你所需要特定主题的一批文献内容。速度快、查询准确，可用秒计算。如需要可随意打印出题录索引。本系统还可按汉语拼音顺序查询，进行主题检索。当用户想查找某一主题文献时，当选择了用汉语拼音途径查询时，屏幕即显示首汉字汉语拼音索引，用户可按所要查找的主题词首汉字进行查找，确定后即可显示此首汉字的所有主题词，从中找到所要查找的某一主题词后，点击鼠标或按回车，即可显示此主题的一批文献。除以上检索途径外，本系统还可从文献专题、文献题名、类别、责任者、期刊题名、模糊查询等多窗口进行查询，以满足用户需要，可打印题录索引。

3.3 《军表》词典库主题词的显示方法

系统无论是从范畴类目进行检索，还是按首汉字拼音途径查询某一主题的文献，都首先要选择主题范围。也就是，首先要系统知道你要查找某一主题词。为了方便用户不必去翻阅《军表》和人工输入主题词，我们采用在屏幕上显示《军表》词典库中的主题词，供用户查询利用。我们所采用的方法是利用菜单显示的方法。当用户进行主题检索时，要通过范畴类目或某一首汉字汉语拼音来查询某一主题词，我们就把该词索引库中的主题词字段赋给一个数据做为菜单数组，当记录行为128个时，再赋给一个数组，以此类推，用户只需翻页就能显示出所需要查询的主题词。确认后，即可检索到此主题的一批文献了。

4 应用效果

我们利用《军表》所开发研制的系统自投入使用以来，深受用户的欢迎，在为教学、科研、学员毕业论文的撰写服务中起到了很好的作用。

1.图书馆开展主题、专题自动化检索服务，进一步挖掘了深层次文献服务，提高了图书馆现代化、自动化服务的层次，开拓了服务领域。

2.本系统自动化程度高，据统计，主题词自动标引率达60%－80%。从手工标引主题词到自动化标引产生了极大的飞跃，提高了主题词标引的准确性，明显地减轻了工作人员的劳动强度，提高了工作效率。

3.本系统提供的主题、专题目录索引和文摘，专指度高，针对性强，提高了文献的查全率和查准率，明显地节约了读者查找文献的时间。

4.功能齐全，检索途径多，用户可从主题、专题、类别、文献题名、责任者、模糊查询等多窗口进行查询检索以满足需要，还可根据读者的需要按类别、专题、主题等打印文献题录索引，方便读者利用图书馆文献。本系统经过专家技术鉴定，给予较高评价，并获军队科学技术进步奖，有推广应用价值。

参考文献：

[1]军用主题词表.军事科学出版社，1990
[2]军用主题词表使用手册.军事科学出版社，1990
[3]刘滨，王源，秦昌.微机辅助文献标引系统设计与研究3.自动标引研究.现代图书情报技术，1997(5)：42－46

国防科技情报系统叙词语言的发展

中国国防科技信息中心图书馆　龚昌明

[摘要]回顾了30多年来国防科技情报系统使用叙词语言——从直接使用国外的主题词表、联合编制书本型叙词表到联合编制电子版叙词表的发展概况，在对《国防科学技术叙词表》(1985年版和1991年版)、《电子版国防科学技术叙词表》(1998年版)进行评述的基础上，简述了国防科技情报系统叙词表的发展成就及今后推进叙词语言研究和应用的若干建议。

[关键词]叙词表　技术评价　检索语言　应用

我国国防科技情报系统早在60年代初期就开始使用叙词语言，并建立了卡片式主题目录检索系统。30多年来走了一条逐步发展的道路：直接使用国外的主题词表→各自编制专业叙词表→联合编制书本型叙词表→联合编制电子版叙词表。这使叙词语言在国防科技情报系统内得到了推广和广泛使用，为我国检索语言的发展和推进文献信息资源的有效利用，起到了重要作用。

1　发展概况

叙词语言在我国的应用较国外晚得多。在我国只是在80年代初由于《汉语主题表》(以下简称《汉表》)的问世才得以广泛应用。然而，在国防科技情报系统，早在60年代初就率先使用叙词语言。1962年8月，原国防科学技术情报资料所为处理引进的西文科技资料(如AD、PB等报告)，建立卡片式主题目录检索系统，采用了《美国武装部队技术情报局主题词典》。这是在我国开创使用叙词语言、建立主题目录检索系统的先例。1971年，航空科学技术情报研究所编成《航空科技资料主题表》，并在航空科技情报系统内推广使用。这是我国第一部自编的叙词表，开创了编制专业叙词表的先例。原国防科委情报研究所在总结近10年使用国外主题词典的经验和存在的问题基础上，于1978年自编成《国防科学技术主题词典》，并在国防科技情报系统内推广和使用。

60年代至70年代，这是叙词语言在国防科技情报系统内的初始使用阶段。

70年代末，国家实施“748”工程，在我国情报界和图书馆界的通力协作下，于1980年编成《汉表》(试用本)。这是我国自编的第一部综合性叙词表。该表的编制是在全国范围内的一次大规模编表实践，它确立了国内自编词表的主流和传统的书本式叙词表的主要编制原则和方法，为我国叙词表的编制奠定了基础。

80年代初，在《汉表》的编制原则和编制技术的基础上，在国防科技情报系统内，电子科技情报所编制了《电子技术汉语主题表》，兵器情报研究所编制了《常规武器专业主题词表》，航空科学技术情报研究所编制了《航空科技资料主题表汉英主题索引》。

虽然，这些专业叙词表发展了一定数量的用户，积累了许多实际的使用经验，但从整个国防科技情报系统的角度来看，多种叙词表的并用，不利于统一的检索系统的建设，更不能有效地实现文献信息资源的共享。为在整个国防科技情报系统内推出一种统一的情报检索语言，1983年，第九

次国防口情报所所长联席会议研究决定，在使用中的4部叙词表的基础上，编制《国防科学技术叙词表》。国防口8个情报单位组成联合编制组，将4部词表进行合并和增补，于1985年编成，并推广使用。这是在国防科技情报系统统一情报检索语言方面的一项重要举措。鉴于该部词表未编制范畴表，中船总为实现本系统各部门科技情报工作的现代化，利用计算机检索情报资料，在《国防科学技术叙词表》(1985年版)的基础上，进行严格的抽词和选词，于1985年5月编制了《国防科学技术叙词表船舶工业范畴表》。航天工业部为使各部属单位完成从分类法向主题法的过渡，于1985年1月成立编辑组，从《国防科学技术叙词表》(1985年版)中选取了与航天专业有关叙词，于1985年6月编成《国防科学技术叙词表航天专业范畴表》。80年代末，鉴于科学技术的发展，经使用多年的专业词表，已不能满足使用要求。有些情报所对其编制的专业叙词表进行了增补和修订，有的新编制了专业叙词表。例如，1988年，中国标准情报中心以ROOT主题词表为蓝本，编制成《标准文献主题词表》；1988年，电子科技情报研究所推出了新版《电子技术叙词表》；1988年，中国核情报中心在全面总结10年来标引实践的基础上，吸收了《INIS叙词表》和《EDS叙词表》的结构特点和选词原则，编制成《核科学技术叙词表》；1989年，北方化工信息中心和兵器工业火炸药情报网新编了《火炸药专业叙词表》；1990年，航空科学技术情报研究所编制出版了《航空科技资料主题表字顺索引和范畴索引》。

从以上简要回顾可以看到，80年代，叙词检索语言在国防科技情报系统得到了飞速发展，在全国进入第一个编表热潮的大环境下，既编有作为统一的情报检索语言在国防口推广使用的叙词表，又有在航天、航空、电子、兵器、船舶、核技术、标准化等专业情报部门推广使用的专业叙词表。对传统的叙词表的结构进行了多样化尝试和创新，并趋于完善和定型。

进入90年代，随着《汉表》修订和《军用主题词表》的编制，全国进入了第二个词表编制热潮。由于《国防科学技术叙词表》已使用近5年，且在编制过程受某些因素的制约而存有一些不尽合理之处。1989年，由原国防科工委情报所牵头，会同航空、航天、电子、兵器情报所、中船总601所和714所、中物院科技情报中心，遵循《军用主题词表编制规则》，对该表进行了全面修订。历时2年，于1991年底编成《国防科学技术叙词表》(1991年版)，并于1992年出版。原国防科工委有关主管部门下发了在国防科技情报系统推广使用该版叙词表的通知，作为国防科技情报系统内统一的检索语言。

虽然，在80年代中期和90年代初期，分别编制了《国防科学技术叙词表》(1985年版)和(1991年版)，拟作为国防科技情报系统统一的叙词检索语言，但由于多种因素，在建的文献数据库仍采用11部叙词表。即：《国防科学技术叙词表》(1991年版)、《核科学技术叙词表》(1988年版)、《航空科技资料主题表》(1990年版)、《电子技术叙词表》(1988年版)、《兵器科学技术叙词表》(1995年版)、《船舶工业叙词表》(1992年版)、《航空航天主题词表》(1993年版)、《航天科学技术叙词表》(1995年版)、《航空航天医学主题词表》(1993年版)、《标准文献主题词表》(1988年版)和《计量学与测试技术主题词表》(1993年版)。这些叙词表的兼容性较差，多种检索语言的同时使用，不利于在国防科技情报系统建立统一的检索系统，已建的各专业文献数据库在联网时难以发挥信息资源共享效益。此外，这些词表多数已使用了多年，已不能反映科学技术的发展，需要增补和修订。但是，传统的书本型叙词表的增补和修订，目前难以获得必要的经费支持。更为重要的是传统的书本型叙词表具有许多固有的不足，如：难以实现与自然语言的转换，改版(增补和修订)周期长，对用户要求高，不能有效地支持计算机辅助主题标引系统，无法在网络环境下使用等。为了满足在"九五"期间国防科技情报电子化工程建设的需要，使自然语言和受控检索语言的优势互补，加速文献数据库的建设，并为实现现代网络化信息服务奠定技术基础，1996年4月，由中国国防科技信息中心牵头，

会同核技术、航空、航天、电子、兵器情报所、中船总综研院和714所、航天总708所、科工委507所、中物院信息中心和中国标准化中心，以《国防科学技术叙词表》(1991年版)为蓝本，联合编制《电子版国防科学技术叙词表》。经过近2年的努力，于1997年12月编成了一部适合国防科技情报系统各部门信息协同处理和网络终端用户信息检索的综合性、动态性电子版叙词表。该部词表1998年投入试用。

2 代表性词表评述

《汉表》是国内编制叙词表的样板，它在国内确立了传统叙词表的主要编制原则和体系结构，首创五表俱全的体系模式。按社会科学和自然科学分卷，各卷又由主表(字顺表)、索引(词族索引、范畴索引、英汉对照索引)组成，并单独将附表列为一卷。主表的显示方式采取：主表的款目均按款目词的汉语拼音逐字母排列，对属分参照采取单级显示，形成字顺表。设立词族索引，以族首词为基本款目，全面揭示具有语义属分关系的同族叙词。以分类原则组织三级范畴索引，所有叙词依据学科为主、词义为辅的原则归入一个或多个范畴。

2.1 《国防科学技术叙词表》(1985年版)

国防科技情报系统联合编制的这部词表，在词表结构和编制技术上进行了大胆革新：它是由字顺表、叙词首字笔画索引和英汉索引组成；未编制范畴表；将词族表并入字顺表，并采取设置领词的技术，建立分词族，使某些词族不要过大；采取了一种全新的显示方式，即等级全显示方式，仅在族首词款目下显示全部等级关系，族中词只作族项参照指引；在词族内配以区间号，据此可较快地在词族中找到某个叙词，并知悉其上下左右是什么叙词；叙词的汉语拼音采取以字为单位标注，不以词为单位连写，从而将相同汉字集中排列。该表由《汉表》的五表俱全模式向新结构演革，作了重大尝试，为国内编制专业表的多样化、趋于完善和定型，起到了有益的促进作用。

该版词表入选正式叙词25502个(其中族首词946个，领词144个)，非正式叙词2481个；正式型号叙词4272个，非正式型号词2262个。总词量为32036个，入口率为9.8%。

2.2 《国防科学技术叙词表》(1991年版)

该版词表遵循《军用主题词表编制规则》，在词表结构和编制技术上继承1985年版词表的特点外，着重对选词、词间关系、英汉索引进行了优化，增编了范畴表，对词表兼容作了先行实践。

对全部叙词的词形作了规范，并增选了与国防科学技术关系密切、文献标引和情报检索所需的名词术语10365个，删去了型号叙词、专有机构名称叙词以及相关学科方面使用频率低、专指性太强的叙词。总收词量为31816个，其中正式叙词28996个(含族首词1006个，领词369个)，非正式叙词2829个。入口率为10.5%。基本上覆盖了国防科学技术领域各学科、各专业及相关专业的知识。

该表进一步优化了叙词词间关系和词形、词义的规范性，增设了参照系统。对全部叙词的英译名进行了规范化处理，编成的英汉索引是一部显示“用”“代”参照关系的简化英文叙词表，可供采用英文叙词建立数据库和英文主题检索系统的单位使用。这种能起简化英文叙词表的英汉索引编制，在国内尚属首次尝试；主表采用全显示方式；叙词款目项中删去了叙词的汉语拼音，增加了英译名，可使主表作为一部汉英词典使用；增编了细分到二级类目的范畴表，并为主表叙词加注了范畴号。为实现与《军用主题词表》兼容，对叙词的用代关系进行了较多的调整。

该部词表在国防科技情报系统和军内发行了1600多套。它的编成,为在国防科技情报系统实现检索语言的统一、文献信息资源的共享奠定了技术基础。

2.3 《电子版国防科学技术叙词表》

这是国防科技情报系统联合编制的第三部叙词表。它融合了国防科技情报系统使用的11部词表,删去了各表原收选的使用频率低的叙词,并增补了大量新叙词。该表不仅是国内投入试用的第一部大型电子版叙词表,而且为融合、兼容化处理不同的专业词表编制大型词表,进行了探索,取得了经验。

为了将11部差异大、体系结构不同的词表的词间关系理顺,将应入选的每个叙词都摆到合适的位置,建成比较科学、正确、完备语义网,编制组提出以《国防科学技术叙词表》(1991年版)为蓝本,采用比对法的编制方法:即其余10部词表收选的叙词,凡经过初审认为应收入电子版叙词表的叙词,应与该蓝本一一比对。在比对过程中,对该表未收选的叙词,采用以下两种方法建立词间关系:

"接枝法"——凡能纳入该表已建词族的叙词,建立词间关系,作为某个或某几个叙词的下位词;

"植入法"——凡不能纳入该表已建词族的叙词,确定其词间关系,建立新词族,作为新的族中词,或作为无关联词。

在编制过程中,编制组着重在以下几方面进行努力:

(1)调整和完备属分关系。对按相关成族原则建立的属分关系、对原先建立的词间属分关系存在跳级的现象、对原先建立的存在差错的属分关系进行了修改;采用双向或多向成族原则建立词族;增收合适的新叙词,使原有的非族中词建立词族;按某些非族中词表述的本质属性,使它们归入已建立的次序;对少数族首词进行了修改;对部分叙词按相关成族和后方一致成族原则建立属分关系。

(2)调整和完备用代关系。对某部词表来说,原先建立的用代关系可能是合理的,具有唯一性。但将11部词表融合在一起,便会出现这部词表确定为叙词,在另一部词表中则可能作为非叙词。为此,对拟入选电子版叙词表的全部叙词和非叙词一一进行了用代关系分析,并对原先建立的不确切、存在差错的用代关系进行了修改;为便于用户使用自然语言向电子版叙词表这种受控语言转换,提高入口率,制订了非叙词的选词原则和采集细则,从同义词、近义词、反义词、通用名和港台名、全称和简称、学名和俗名、外来语音译名和意译名、英汉混合词与全汉语词、现用名与曾用名、同义异形词与曾用过的简体词、词素相同构词手段或辅助构词手段不同的名词术语等11方面,建立用代关系;此外,建立了词间组代关系。对某些概念复合度过高、可采用2个已入选的叙词组配表达的复合概念叙词,建立了组代关系。从而使非叙词的收词量达22405个,入口率提高到43%。一个叙词的代项最多达8个。

(3)完备参照关系。书本型叙词表受篇幅的限制,建立的参照关系一般不能太多。电子版叙词表可充分利用接收机内存量大的有利条件,建立比较完备的参照关系。据此,对各表原建立的参照关系进行了增删。现每个叙词的参照项,平均超过2个,最多的一个叙词的参照项达60个。

(4)增加含义注释。对有些仅从词面上不易正确理解其表述的概念的外延和内涵的叙词,增加了注释。

(5)补加了词义限定。对有些词形相同、在不同专业其概念外延和内涵有所不同的叙词,补加了词义限定。

(6)对叙词英译名进一步作了规范。对入选电子版叙词表的每一个叙词的英译名进行了规范化处理,使每一个叙词都有对应的英译名,可数名词采用复数;当一个叙词有2个不同的英译名时,根据习惯用法,选定其中一个译名为英文叙词,作为第一英译名,可用于标引文献,另一个译名为非英文叙词,作为第二英译名。经过对叙词英译名的规范化处理,保证了该版叙词表与国外著名叙词表的兼容,并为直接从英文叙词进行文献标引和检索,提供了技术手段。

(7)参考了国内外多部著名的叙词表和多部权威性工具书,使编成的电子版叙词表收词量大(总收词80005个)、专业覆盖面广、收选了大量反映新技术新概念的叙词、词间关系准确完备、语义齐全、叙词的英译名规范。

电子版叙词表不是传统的书本型叙词表简单的机读版,必须能在网络环境下运行,适用于计算机辅助主题标引作业。为此,在机内建了一个主库和几个辅库,如属分关系辅库、用代关系辅库、参见关系辅库、自由词辅库等。通过专用程序,将它们有机连成一体,形成电子版词表。

为便于文献主题标引和信息检索时使用,该电子版叙词表还设有实时使用模块,由三个子模块组成:

(1)叙词导航模块;它为文献主题标引人员和终端信息检索人员提供8种查词手段,实现将文献主题概念或用户的检索提问转换成规范的叙词。

(2)叙词定位模块;主要实现词族查询和基于经验的叙词联想功能。

(3)叙词范畴选择模块;主要依据文献标引人员给出的叙词,自动将文献归入范畴。

为便于电子版叙词表及时更新和维护,还建有一个后期维护模块。主要用于标引(检索)信息、使用过的自由词及其信息的自动保存、词表的动态管理。

该电子版叙词表已在中国国防科技信息中心等单位的网络上运行,在国内首先实现检索语言的电子化和实用化。

3 对推进叙词检索语言研究和发展的若干建议

当前,我国国防科研体制发生了重大变化。在此新形势下,如何加强对国防科技情报系统检索语言研究和发展的领导,应采取何种对策,以便推进国防科技情报系统文献数据库的建设和信息资源的有效利用?

在叙词检索语言研究和发展方面,过去30多年,国防科技情报系统走出了一条在上级主管部门领导下联合研究和发展的道路。这具有中国特色,适合我国国情。在国防科研体制发生了重大变化的新形势下,现必须寻求一条推进国防科技情报系统叙词检索语言研究和发展的新机制。在上级主管部门领导下采取联合研究和发展,仍是一条走向成功之路。鉴于各个国防工业情报部门归属了国务院国防工业委员会,多个军队研究所归属了总装备部,建议发挥中国国防科技信息学会的作用,在上级有关主管部门的领导下,由中国国防科技信息学会组织有关国防科技情报部门,组建“国防科技情报系统检索语言管理委员会”这一协作性机构,承担叙词检索语言联合研究、推广使用的组织、协调工作。

叙词检索语言的研究和电子版国防科学技术叙词表的编制,关系到国防科技情报系统数据库的建设和信息资源的共享及其有效利用,是实现现代化网络化文献信息服务的基础性建设。鉴于全文检索系统在查准查全率方面尚一时难以解决,应开展受控检索语言与全文检索系统优势互补的研究。建议有关主管部门将它列为国防科技情报电子化工程的子项目,每年安排一定的专项经费予以支持。

国防科技系统现已从联合编制书本型叙词表进而在国内率先编制和使用电子版叙词表。随着一体化文献处理系统的推出，应加强在国防科技情报系统，特别是在总装备部系统中推广和使用，举办学习班，从而获取使用经验，发现不足，使其更加完善。

《电子版国防科学技术叙词表》作为一种受控检索语言，与传统的书本型叙词表一样，滞后科学技术的发展。为使其具有生命力，必须建立一种修订和改版机制，并配备专职人员，跟踪科学技术的发展，积累新出现的反映新学科新技术的名词术语。为此，建议在"国防科技情报系统叙词检索管理委员会"下设立"电子版国防科学技术叙词表联合管理办公室"这一非编协作单位，由有关情报部门主管叙词检索语言研究和叙词表编制工作的专业人员参加。其任务是负责搜集本系统使用《电子版国防科学技术表》的意见，跟踪本系统新学科新技术的发展，积累新出现的反映新学科新技术的名词术语，定期将修改和增补意见提供给词表联合管理办公室，作为词表每年修订和改版一次的依据。

我国当前国防科技情报研究已涉及政治、外交、军事、经济、科研管理等领域。但是，《电子版国防科学技术叙词表》基本上是11部技术性叙词表的融合体，收选政治、外交、军事、经济、科研管理等方面的叙词，明显不足；入口率尚需提高，以增强电子版叙词表与自然语言的自动转换能力，满足当前情报研究和数据库建设的需要。

此外，国防科技情报系统从事叙词检索语言研究和词表编制工作的专职人员明显不足，在这方面具有专长的老同志大都将退休。建议有关部门应调配合适的人员，并重视中青年检索语言研究和词表编制专职人员的培养，以便叙词语言的研究和开发后继有人。

参考文献：

[1]龚昌明.电子版国防科学技术叙词表编制技术报告,1997

[2]叶千军,朱丽.试论80年代以来我国叙词表的发展——兼评《军队档案常用主题词表》.图书情报工作,1997(2):16—20

建立通信兵文献标引检索语言　实现通信兵文献管理标准化

解放军理工大学图书馆　彭延生　田毅农

[摘要]我馆在继编制《军用主题词表》之后，近两年来又组织了《通信兵主题词表》的编制。本文概要介绍了《通信兵主题词表》编制的意义，结构体系，编制特点及编制体会。旨在为信息管理人员使用、编制词来提供参考、借鉴。

[关键词]通信兵　主题词表　词表编制

总参通信部于1999年1月在北京召开了《通信兵主题词表》(下称《通表》)的审查会，通过了对国家军用标准《通信兵主题词表》的审查。《通表》不久将作为国家军用标准在全军通信部队颁发实施。这标志着我军通信兵文献信息管理走向了标准化。《通表》的建立，为建设通信兵文献信息资源共享奠定了基础。

1　《通表》是通信兵信息化建设的基础

当前，科技发展日新月异，知识经济初见端倪，世界军事革命方兴未艾，军事通信技术也正在以前所未有的速度发展。现代化的信息技术要求具有高度的自动化程度和信息处理速度，它不但依赖先进的硬件设备，还需要许多与之相配套的"软件"，需要对信息进行标准化处理。主题法理论的应用和发展，适应了这一趋势。编制主题词表是军队信息化建设的需要，也是通信兵信息化建设的需要。军委江主席指出"四个现代化，哪一化也离不开信息化"。因此，建立《通表》是与我军现代化建设紧密相关的。

《军用主题词表》是一部综合性的词表，所收的通信方面的主题词有限，难以满足通信兵各专业领域的需要。目前，在军事通信、指挥自动化、计算机技术、无线电管理等领域，涌现出大量的新概念、新术语。特别是今天因特网已遍布全球各个角落，我军业已建成了军事训练信息网，实现了军事指挥、训练的网络化和自动化。为充分发挥网络的功能，迫切需要制出一部具有科学性、权威性、有约束力的文献信息管理的主题词表。

《通表》是《军用主题词表》系列的重要组成部分，是我军通信兵第一部大型的专业性主题词表。《通表》与相关的国家标准和国家军用标准协调一致，结构合理完整，规范化程度高，突出了专业特色，具有科学性、实用性，达到了当前国内外同类词表的先进水平。

2　《通表》的结构体系

《通表》由前言、编制说明、字顺表、范畴表、汉英－英汉对照表和附加说明组成。字顺表的主题词，除少量无关词外，均通过用项(符号Y)、代项(符号D)、属项(符号S)、分项(符号F)、参项(符号C)等手段建立了相互联系。其中，"用项"用于指引正式主题词，"代项"用于指引非正式主题词，

“属项”用于指引直接上位主题词，“分项”用于指引直接下位词，“参项”用于指出与其相关的主题词。字顺表排列方式借鉴了现代汉语字典的排序方式，符合传统查字习惯，便于使用。

范畴表是为方便按分类查找主题词而编制的，它是进行文献标引与检索的辅助工具。范畴表中的主题词按词义和学科属性分别归入各自的范畴，对于少量其概念被两个以上学科作为重点研究对象的主题词，则同时归入不同的范畴。《通表》范畴表设一级、二级范畴，一级范畴按军事通信系统职能关系的顺序共设法19个范畴（通信指挥；通信联络组织；通信保障；指挥自动化；通信兵战备；通信装置；通信兵训练；无线电管理；通信科研；通信工程；通信台站勤务；通信原理；信息技术；计算机技术；通信技术；通信基本电路；常用军事术语；相关学科；共用语），二级范畴按“从一般到具体”的顺序共设128个，用阿拉伯数字加汉语拼音字母的方式来表示。

汉英—英汉对照表是为便于查找主题词英文译名而设置的，是词表的辅助查询工具。该表由汉英对照索引和英汉对照索引两部分组成。汉英对照索引中的主题词，按字顺表主题词首位汉字拼音排序方法排列，汉语主题词在前，英译名在后。首字非汉字的主题词，以阿拉伯数字、希腊文、俄文的顺序排列。英汉对照索引中的主题，是按字顺表的主题词逐一比对英文字母的排序方法，英译名在前，汉语主题词在后。英译名中的空格、连字符等符号按标准排序。

3 《通表》的编制特点

在《通表》编制过程中，始终坚持“质量第一，方便使用，适当超前”的指导思想，严格遵循国际与国家的相关标准，强调突出我军通信兵特色，认真吸取国内外有关词表的精华，力求编出一部科学性与实用性统一，综合性与专业性兼顾，主题词明细度适当，参照系统完备，汉英—英汉对照准确的专业主题词表。

《通表》具有以下六方面的编制特点：

3.1 编制技术方面

《通表》在编排方式上采用了国际国内流行的编表方式，版式按照国家军用使用标准样型设计，词表版面采用双幅式排列方式；索引采用了首汉字汉语拼音索引和首汉字笔画索引方式两种方式。字顺表主题词的编排采用了全显示方式，主题词用、代、属、分、参款目格式齐全，词间关系符号替代性强，词与词之间联系紧密。范畴表中的主题词，通常按其词义和学科属性归入一个范畴；若某一主题词所代表的概念被两个以上学科作为重点研究对象时，则同时归入不同的范畴内。范畴的等级采用层累制，前两位阿拉伯数字表示一级范畴；后一位汉语拼音大写字母表示为二级范畴（为避免数字与汉语拼音字母混淆，省略字母“I”、“O”）。汉英—英汉对照表中确定了英译名主要为名词或名词性词组等15项英译规则。

3.2 收词范围方面

《通表》选收的主题词主要来源于四个方面：即1990年版、1997年版《中国军事百科全书》军事通信技术部分；1990年版《军用主题词表》本专业主题词及相关词条；各类军事通信条令、条例；国内外通信专业有关文献。记录显示，表内共收录主题词14928个。其中，正式词8942个，非正式词5986个。族系中，族首词582个，领词82个。经辨析，表中所选定的主题词均为军事通信领域16个专业门类文献中经常出现的名词术语，单个主题词字数一般控制在12个汉字以内。

3.3 用词表述方面

《通表》选定的主题词词形、词义规范，符合选词规则，组配功能较强，使用频率较高。词表中的词汇控制水平比较高，经比对，非正式主题词与正式主题词之比即入口率约为67%，非正式主题词指用代基本恰当。词表中的受控语言规范，无关联词少，先组度低。

3.4 实际运用方面

经有关单位试标引，共标引文献780篇，使用标引词3156条次，其中，表内主题词占标引量的95%，未收的主题词161条次，占5%。

3.5 编表自动化程度方面

《通表》使用VBA开发工具设制的Windows 95平台上的《通表》编制管理系统，界面友好，运行稳定；可对词条进行增、删、改等操作，具备编辑、校对、查询、统计等8种功能，并能在屏幕上显示主题词的用代关系、属分关系、相关关系、族系关系和释文等，输出打印格式可根据编表需要自行设置；系统中的遇错自动告警功能为使用者提供了便利条件。

3.6 数据库系统方面

《通表》在技术运用和保障上建有基础库形式的数据库，库中每一主题词都用阿拉伯数字定位编号，组表时可实时地、动态地形成不同应用范围的4种表，即字顺表、范畴表、汉英对照表、英汉对照表。为保证主题词定位，每一主题词在生成过程中，必须5个要素齐全，即编号、拼音、范畴号、主题词条、英译名。同时还配有该词的释义，用以检验该主题词是否准确。

4 编制《通表》的经验体会

4.1 抓好三个重点环节，确保了编表的质量

牢固树立面向用户的思想，发挥集体的智慧和力量。在编制《通表》工作中，重点抓了选词、词控和英汉互译三个方面的工作。

(1)在选词方面注意突出一个"新"字。针对科学技术日新月异，在普遍选词的前提下，下功夫选择新的主题范围，确定新的主题词。编制人员充分利用图书馆文献、资料丰富的得天独厚的条件，参考了大量的中外期刊、科技专著、军内外相关的主题词表和新版辞书。经过三次审稿四次修改。同时又增加了"信息技术"范畴，增添新词近千条。为了保证这些新选主题词的质量，充分利用我院通信、指挥自动化学科门类齐全，研究水平较高的条件，邀请各专业的专家共同研讨、审定。

(2)在提高词控水平方面，注意突出一个"广"字。词表的词控水平是衡量一部词表质量高低的重要标志，其主要表现在入口率和后组度高低两个方面。为提高词表的词汇控制水平，实现《通表》提出的入口率达到50%以上的要求，集中人员、集中时间对《通表》中的主题词逐一审核，参阅军内外相关的主题词表、辞书、军语、大百科等资料，大大扩展了主题词的入口词的范围。《通表》的19个范畴中除"相关学科"外，都达到60%以上的指标，其中"09 通信技术"范畴入口高达89%。

(3)在主题词英汉互译方面，注意突出一个"准"字。英汉对照表是《通表》的重要组成部分，是文献索引、引咨外文资料的重要工具，英译名正确与否，会影响到《通表》的权威性和规范性。为保

证英汉互译的准确性，认真引用国际权威的工具书，吸取了专家、教授对英译的意见，进行了四次比较大的修改，最后使汉英—英汉对照表基本实现了标准互译。

4.2 为了保证《通表》的科学性和实用性，在完成《通表》送审稿基础上，必须进行试标引

在试标引工作中，注意了三点：

(1)在参加试标引人员的选择上，注意代表性。请机关部门、专业系的公文管理人员、图书资料编目人员以及文献资料咨询检索人员，模拟《通表》用户，参与试标引工作。

(2)在试标文献的选择上，注意广泛性。我们从不同角度、不同领域，选择了公文、电子及通信类科技期刊、图书、研究报告和学术论文进行标引。

(3)在试标引方法上，注意多样性。按照标引规则，对公文、图书采用浅标引方式，对文献资料及期刊采用深标引方式。

《通表》是全军通信兵系统基础性建设的一个重要项目，也是军事通信领域文献检索语言重要科研课题。《通表》通过审定后，我馆又开始推广使用，并着手释义词典和光盘配套工程的编制工作。

论我国文献分类法产生和发展的基本规律

郑州大学信息管理系　刘廷章
铁道部郑州公安管理干部学院图书馆　苗　琳

[摘 要]文章阐述了中国文献分类法源于书目分类体系、深受外国分类法的影响、随着科学技术的发展而发展、随着社会的变革而变革、以用户的需求为转移、随着自身的修订而发展这六条基本规律,同时就涉及到的一些有争议的问题谈了作者的看法。

[关键词]文献分类法　基本规律　中国

前言

我们知道,中国文献分类法这条河源远流长。所谓源远,可以追溯到公元前26年刘向、刘歆的《七略》,就像长江可以追溯到沱沱河一样。但源不等于流,沱沱河是长江的发源地,但不等于长江;《七略》是我国文献分类法的起源,但决不是中国第一部文献分类法。所谓流长,中国的文献分类法如果从1917年沈祖荣、胡庆生的《仿杜法》算起,至今已有83年的历史了,而且它会像汇纳百川而入东海的长江一样,继承书目分类体系的优良传统,借鉴国外分类法的成功经验和先进技术方法,随着科学技术和社会发展的日新月异,根据用户日益增长的需要不断地修正自己,发展自己,以至于无穷。本文拟就中国文献分类法在其产生和发展中的基本规律及涉及到的争议和问题,谈谈自己的看法,以求赐教。

中国文献分类法在其产生和发展的历史长河中,有其自身的基本规律,概括起来主要有以下几点:

1　它起源于书目分类体系

我国的文献分类法与外国的文献分类法一样,都是起源于书目分类体系的。这是大家所公认的一条客观规律。问题在于书目分类体系与文献分类法到底有没有区别?(首先声明一点,书目其实也属于文献的范畴,但为了便于研究,我们应该把二者分开来看,这在我国历史上已经早已分开了,比如书目属于目录学的研究对象,而文献则属于文献学的研究对象。)笔者认为二者是有本质不同的,其主要区别就在于:书目分类体系是依赖于和依附于具体书目,不同的书目都有与众不同的分类体系,它是供检索具体书目而用的,相当于一书中的"目次",它属于中国目录学的研究内容。而文献分类法是独立成册的文献分类工具,所有的文献都可采用同一个分类体系,它是供检索所有文献信息而用的,相当于知识库中的一把"钥匙",它属于中国文献分类学的研究内容[1]。

有人有意无意地混淆了上述区别,从而得出了不符合客观实际的结论。比如,目前在我国文献分类学界似乎有一种定论:《七略》所反映的分类体系,就是我国最早的一部文献分类法。从《七略》

到《四库》(《四库全书总目》之简称)这段历史,就是文献分类法发展史[2]。那么,如何衡量一部书是书目还是文献分类法呢?笔者认为主要有三条标准:第一,是否有完整的类目体系;第二,是否有完善的标记符号制度;第三,是否独立成册成为类分所有文献的分类工具[3]。由此可见,《七略》是我国第一部"部次群书"的"中国汉代官府藏书目录"。"《七略》中创立了以儒家思想为指导的书目分类体系,为中国目录学的发展奠定了基础"。它所反映的分类体系是书目分类体系,而不是我国最早的文献分类法。我国最早的文献分类法应首推沈祖荣、胡庆生合编的《仿杜威书目十类法》。同样,《四库》是"中国清代官修书目",而不是比较完善的中国图书分类法。因此,从《七略》到《四库》这段历史,决不是什么中国文献分类法的发展史,而是书目分类体系的发展史。《图书馆学百科全书》在"条目分类目录"中也是这样划分的[4]。《仿杜法》很多类目设置,如"000 经部、类书"等,借鉴了《四库》的分类体系,就是说《中图法》也可以容纳经、史、子、集所包含的古籍,这不能不说《中图法》在某些方面继承和借鉴了《四库》的分类体系。因此说,我国的文献分类法是起源于书目分类体系的。

2 它深受国外分类法的影响

从1917年到1951年这35年间,我国编制了60多种分类法。这些分类法都是以DDC为蓝本的,并因此分为三大派别:仿杜、改杜和补杜。

从1954年的《人大法》到1975年的《中图法》这22年间,我国共编制了28部分类法。这些分类法在体系结构、标记符号、标记制度等许多方面,都深受ББК和前苏联其他分类法的影响。如《人大法》初稿在"图书分类法编制的简略说明"中谈到了"学习苏联托罗帕甫斯基著作的苏联大众图书馆适用分类法和克列诺夫著作的苏联小型图书馆适用的分类法简表"的问题。从大类体系来看,《人大法》与托罗帕甫斯基的《十进分类法》颇有相同和相似之处。又如《科图法》在第一版说明中也谈到了"我们参考了'苏联图书分类法草案'基本大类和安巴祖勉等编的图书分类法"。因此,《科图法》在许多方面都打上了"苏联图书分类法草案"的烙印。从《中小型表》到《大型法》再到《中图法》,它们是一脉相承的。由于《中小型表》在许多方面都受到了ББК的深刻影响,那么,作为一脉相承的《中图法》也必然带有ББК的印记。如关于类组的设立、采用多重列类方法、采用混合号码、增设专用复分表及编制系列版本等方面,都借鉴了ББК的成功经验,当然作为广采众家之长于一身的《中图法》,也借鉴了CC、UDC、DDC等许多国外分类法的先进技术和科学方法,如CC的八分法、小数标记制、组配法等,UDC的辅助符号、双位制、借号法等,DDC的双表列类法等。

从1917年的《仿杜法》到1975年的《中图法》,中国的文献分类法经历了从模仿到创造并逐步走上成熟的过程。其中有三大里程碑:一是《仿杜法》,它是我国第一部从书目实体中脱颖而出的分类工具,是"第一个为中文书而编的新型分类法"(杜定友语)。二是《人大法》,它是我国第一部具有中国特色的社会主义分类法,它"开始了我国图书分类法史上一个新阶段"(刘国钧语)。三是《中图法》,它是我国第一部全国通用的大型综合性的文献分类法,它开创了我国文献分类工作走上三化(标准化、自动化、分类主题一体化)的新纪元。但不管是哪一个阶段,哪一种分类法,在我国文献分类法产生和发展的历史长河中,我国的分类法都受到并将继续受到国外分类法的深刻影响。有人有意无意地否认这种影响,只承认继承不承认借鉴,这是不符合我国文献分类法产生和发展的客观规律的。

3 它随着科学技术的发展而发展

1917年印行的《仿杜法》,在10个大类里,只列出了"500 科学"1个大类;1953年问世的《人大法》,在17个大类里只列出了4个大类;1957年出版的《中小型表》,在21个大类里自然科学占了7个大类;1963年出版的《大型法》,在《中小型表》的基础上发展为8个大类;1975年出版的《中图法》,在《大型法》的基础上增加到10个大类。由于科学技术的发展,迫使文献分类法不得不增加新的类目,用以容纳新产生的科学技术文献。如在20世纪60年代产生、70年代才得以发展的环境科学[6],在《中小型表》和《大型法》里没有反映,而在《中图法》里作为"X 环境科学"一个基本大类独立列出。人类社会进人80年代之后,以信息科学、生命科学和材料科学为前沿,以电子计算机、生物工程、空间技术、海洋开发、光纤通信、激光技术、新材料技术、新能源技术等八种新技术为先导的新技术革命,正在全球兴起,我国也实施了"863 计划"[7],由此而产生了大量文献,这在《中小型表》和《大型法》里是不可能得到反映的,而只有在《中图法》里才能得以体现,特别是在《中图法》第四版体现得更充分。

由上述可知,我国的文献分类法是随着科学技术的发展而发展的,这也是大家所公认的一条客观规律。问题在于,文献分类体系与科学分类体系有没有区别?笔者认为,二者是有本质不同的。其主要区别就在于:科学分类体系是根据各门学科的基本特征及其学科之间的联系与区别,确定每门学科在科学体系中的地位,它是自然辩证法研究的重要课题;文献分类体系是根据文献的内容特征和形式特征概括为众多类目,然后按照类目之间的联系与区别编排为一个分类系统,它是文献分类学研究的重要内容。有人有意无意混淆了这种区别,从而得出不符合客观实际的结论。比如,目前在我国文献分类学界似乎有一种定论:文献分类体系必须以科学分类体系为基础,文献分类体系必须与科学分类体系一致起来。那么,客观实际又是怎么一回事呢?就拿人们公认为以科学体系为基础的《中图法》来说,第一部类和第五部类显然不能说是以科学体系为基础的,就拿与科学体系关系密切的"哲学"、"社会科学"、"自然科学"三大部类的分类体系来说,也是与科学分类体系大相径庭的。当然,也有以科学分类体系为基础的,如《中图法》中化学类、矿业工程类、语言类(除汉语、中国少数民族语言和常用外国语外),但都是失败的。这是因为,文献分类法是文献分类的工具而不是科学分类的武器,它有其自身的特点[8]。

4 它随着社会的变革而变革

在半殖民地、半封建社会形态下的中国分类法,不可能不打上当时社会形态的烙印,如许多分类法都把经学和宗教列为基本大类,并把经学冠于首类[9],这除了受DDC的影响之外,最主要的还是分类法编制者受到当时社会的局限。建国以后的50年,随着社会的变革,人们的思想认识也产生了质的飞跃,同样在分类法中也充分体现了这种变化。解放以后所有的分类法,几乎都把"马列主义、毛泽东思想"单独立类并冠于首位,而把经学和宗教降为小类列出。虽然说是受到了ББК的影响,但最主要的还是社会的巨大变革使然。人们往往批评《中图法》的社会科学部类不如《科图法》,尤其是《中图法》把"政治"放在"经济"大类之前,并把"法律"放在"政治"之后,合并为一个大类组,使人难以理解。如果把分类法与产生它的时代联系起来,就不难发现,分类法是随着社会的变革而变革,发展而发展的。《科图法》产生于国民经济恢复和发展时期,当时强调经济是基础、上层建筑为经济基础服务。而《中图法》则产生于"文化大革命"时期,当时强调政治是统帅,是灵魂,

砸烂"公、检、法"，如是才有上述两种分类法的两种结果。过去《中图法》、《科图法》按政治观点区分类目、以前的《人大法》按国家性质划分类目，现在看来都是不可思议的，但在当时又是合情合理的。再如，社会学、管理学、人口学等在过去把它们视为禁区，改革开放的社会变革，才使人们认识到它们都是社会科学中的重要门类，并在分类法的修订中得到了充分的反映。总之，社会的变革必然要反映到分类法中，反之，分类法也必须随着社会的变革而变革，这可以说是分类法发展的又一条规律。

目前我国现行的三大分类法(此指《中图法》、《科图法》、《人大法》)，虽然都经过了多次修订，力求随着社会的变革而变革，跟上时代步伐，但总的来看仍显得滞后。当然，人们对于《中图法》历史上乱贴政治标签的现象是记忆犹新的，如对五大部的第一部，仍感到不动比动保险，特别是 1999 年出版的第四版，要在 A 大类增加邓小平理论，实在是失策。对于市场经济、知识经济、网络经济所带来的社会变化和信息浪潮的冲击，三大分类法中很少反映，这些都会影响到分类法的生存和发展。

5 它必须及时修订才能生存和发展

1957 年由中国政府组织编制并颁布的《中小型表》，从理论基础到编制技术乃至人力、物力等方面，都优越于 1954 年出版的《人大法》和 1958 年问世的《科图法》，曾经得到过中国许多中小型图书馆的青睐，但为何 70 年代之后再也无人问津了呢？问题就在于 1966 年由北京图书馆修订 2 版之后，再也没有专门机构去管理它了。《人大法》于 1953 年出版，1954 年出增订第 2 版，1957 年出增订第 3 版，1962 年出第 4 版，1982 年出第 5 版，1996 年出第 6 版，平均修订周期是 8.6 年。但因从 1982 年到 1996 年 14 年里只修订了 1 次，即未能及时修订，并因修订质量不高，导致多数用户改用《中图法》。《科图法》于 1958 年出版，1979 年出修订第 2 版，1994 年出修订第 3 版，平均修订周期是 18 年。但因修订周期过长，特别是从 1979 到 1994 年 15 年里只修订了 1 次，也是未及时修订，导致了许多用户改用《中图法》。《中图法》从 1975 年出版之后，1980 年出第 2 版，1990 年出第 3 版，1999 年出第 4 版，其修订周期平均为 8 年。由于《中图法》能及时修订，而且修订质量高，受到用户普遍欢迎，1990 年在全国图书馆界曾掀起了一个学习和推广《中图法》的"三版热"，被誉为"《中图法》年"[10]。由此可见，任何一部分类法，其编辑的终结，也就是修订的开始；不修订就会寿终正寝；修正不及时，也会日落西山。

从上述可知，及时修订是分类法得以生存和发展的关键所在，这是大家所公认的一条客观规律。目前的问题主要有三：一是预见性差，修订质量不高。《人大法》几次修订，对新兴学科增补有限，对高新科技缺乏预见性，该法总是老气横秋，导致大量新学科、高新技术的文献无"家"可归，许多用户不得不改换"门庭"。二是修订幅度掌握不好，用户不满意。《中图法》第一次修订时提出分类法要为现实政治斗争服务，增补了许多不稳定性的类目，第二次修订时不得不进行大的调整和删改。而第三次修订时，采取了"一般不作轻易的变动"的原则，致使广大用户提出的问题没有完全解决。违背了"上帝"的意愿。三是修订周期太长，时差大。前面已述，《人大法》从第 5 版到第 6 版，经过了 14 年，《科图法》从第 2 版到第 3 版经过了 15 年，《中图法》从 1983 年开始修订，直到 1990 年才得以出修订版。而科学技术从 20 世纪 50 年代以后，每隔 10 年就翻一番，到 70 年代，每隔 5 年增加 1 倍，进入 80 年代后，大约 3 年就要增加一倍，90 年代，在 1 年半就翻一番。作为随着科学技术发展而发展的文献分类法而言，就不得不随时修订，及时跟上科学技术发展的步伐。

6 它以用户的需求为转移

作为以类分文献为己任的文献分类法必须以用户的需求为转移,这恐怕是一条非常重要的规律。经过历史的筛选,现在我国还在使用的主要有前述三家分类法,而在这三家分类法中又以《中图法》为最,全国已有94%的单位采用了它。除了它有系列版本、修订及时、宣传有力之外,主要在于它类目详细,注释较多;能充分反映当代的新科学、新技术;编制技术先进,如采用组配编号法、多重列类法、双表列类法、突出列类法等;分类号及配号方法简短易记等。总之,它的实用性强,符合广大用户的需求,因此受到了普遍的欢迎。它还能及时修订,紧跟科学技术的发展,尽可能满足用户的需要,如用户早就迫切需要分类自动化,而《中图法》第四版就增加了机读版。当然它也有很多问题,主要是:(1)人为性的加"0";(2)复分太复杂;(3)未设综合性学科大类;(4)体例不统一;(5)同位类外延交叉多。这些问题都曾经并一直受到广大用户的抨击。尽管《中图法》编委会设置"用户联系卡",召开座谈会,征集批评文章,经过多次修订,但遗憾的是《中图法》编委会对此并未作出深刻的反思。因此,有人提出为21世纪编制一部新型的分类法,实现分类法"由繁及简的再一次飞跃[11]"。

7 几点结论

综上所述,可以得出以下几点结论:一、我国的文献分类法,不是书目分类体系的延续,也不是科学分类体系的翻版,更不是外国分类法的复制,也不是政治斗争的工具,而是继承、吸收、借鉴众长于一身的适合于图书馆使用的根本大法。二、我国文献分类法应在继承书目分类体系的同时发展自己,在借鉴外国分类法的同时创新自己,在吸收科学分类成果的同时壮大自己,在调查用户的基础上修订自己,在修订的同时完善自己。三、创造简便、动态、实用、符合国情的新型分类法,将是21世纪我国文献分类学研究的重大课题。

参考文献:

[1]刘延章. 文献信息分类学. 北京:中国科学技术出版社,1996,4:9-11
[2]吴仲强. 中国图书馆学史. 长沙:湖南出版社,1991,12:154-161
[3]刘延章. 文献分类学. 郑州:中州古籍出版社,1990,3:304-319
[4]《图书馆学百科全书》编委会. 图书馆学百科全书. 北京:中国大百科全书出版社,1993,8:36-37
[5]刘延章. 中国当代分类法对外国分类法的借鉴和在借鉴中的得失. 图书情报工作,1992,(2):17-21
[6]潘永祥等. 自然科学概述. 北京:北京大学出版社,1986,11:481-495
[7]刘炳延. 当代高新科技社会学. 长沙:湖南科学技术出版社,1997,7:60-73
[8]刘延章. 是科学分类,还是文献分类? 图书情报论坛,1990,(2):8-11
[9]北京大学图书馆学情报学系《图书分类》编写组. 图书分类. 北京:书目文献出版社,1990,9:233
[10]刘延章. 分类法修订散论. 晋图学刊,1992,(2):16-19
[11]叶常青. 对当代图书分类法发展趋势的思考. 全国分类法、叙词表发展方向学术研讨与成果展示会论文34号

《中国图书馆分类法》发展的历史回顾与展望

国家图书馆　汪东波　卜书庆

[摘要]本文回顾了《中图法》的发展历史，着重介绍了《中图法》第四版的修订指导思想和修订原则，总结了《中图法》发展的成功经验，展望了《中图法》今后的发展趋势。

[关键词]《中图法》　历史　趋势

1　《中图法》发展的历史回顾

《中国图书馆分类法》(以下简称《中图法》)是我国目前图书情报界广为使用的一部综合性分类法。

《中图法》的前身可追溯到1957年由文化部社会文化事业管理局公布的《中小型图书馆分类表草案》，草案所确立的"五分法"基本体系和混合制标记符号直接为《中图法》所继承。1959年举办的"全国省市图书馆馆长进修班"上集中研讨了编制大中型与专业图书馆用分类表问题。随后，在中央文化部和教育部的主持下，由北京图书馆牵头组成了图书分类法编辑组，着手编制《中国图书馆图书分类法》(后俗称《大型法》)。1964年图书分类法编辑组内部出版了《中国图书馆图书分类法草案》(下册)，下册中包括自然科学和附表部分。1966年3月初又油印公布了其上册的未定稿，内容包括哲学和社会学科部分。由于"文化大革命"的原因，草案与未定草稿都没有最终的结果。《大型法》虽然是一部未完成的图书分类法，但其体系结构，标记制度以及编表技术为《中图法》所借鉴。

1971年2月，在国家文物事业管理局(国务院原图博口领导小组)的关怀和支持下，北京图书馆倡议以大协作的方式编辑一部新的图书分类法，这一倡议随即得到全国各系统图书馆的积极响应，先后参加编表工作的有省、市、自治区图书馆、高校图书馆以及中国科技情报所的36个单位，经过两年多时间的努力，于1973年3月完成草表并由北京图书馆以试用本的形式印出。1974年7月至11月，由编辑组在京6个单位共同参加，在广泛征求各地图书馆意见的基础上，对该试用本进行了修订、补充，于1975年10月由科学技术文献出版社正式出版，产生了《中图法》的第一版。

《中图法》1975年出版后，陆续为全国许多图书馆和情报文献单位所采用，它较好地解决了大型图书馆的图书分类问题。但是由于当时的客观条件所限，特别是当时很大程度上受政治形势的影响，致使分类表中出现了相当多的政治口号乃至在后来可视为政治性错误的问题。鉴于分类表中的这些情况，1979年在长沙召开了有关《中图法》修订的工作会议。会上确定了修订方针、原则和修订重点，并成立了经国家文物事业管理局批准的《中国图书馆图书分类法》编辑委员会(简称《中图法》编委会)，以接替原来的《中图法》编辑组，负责对《中图法》第一版进行修订。第一版修订工作始于1979年4月，具体由《中图法》编委会下设的《中图法》修订组承担，修订组用4个月时间完成修订稿，后经《中图法》编委会全体会议审定通过，由书目文献出版社于1980年6月出版，此为

《中图法》的第二版。

1980年12月在南宁召开了"全国分类法、主题法检索体系标准化会议",会上,全国文献标准化技术委员会第五分会建议"以《中图法》第二版为基础,通过修订完善、充实提高,以作为国家标准分类法",国家标准总局采纳这一建议,以国标发字304号文通知有关单位。从1983年开始,《中图法》编委会着手分阶段地对《中图法》第二版进行修订。1984年11月在南京召开了《中图法》编委会扩大会议,会上研究并决定了《中图法》第二版修订方针、原则、重点、方法、步骤,随后成立了55个专业修订小组,至1990年2月由书目文献出版社正式出版,该版即为《中图法》第三版。

《中图法》自第三版定稿至1996年已有8年多时间。社会发展、科技发展带来了文献出版情况的变化,我国图书情报界计算机编目和检索已相当普及,各种数据库大量制作,区域联网和国际联网得到了应用,由此对《中图法》产生了影响,提出了新的要求,包括对分类法的功能,修订出版形式,修订方式,修订改版后文献的处理方法,对新学科、新主题以及各种类型出版物和网上信息归类的实用性问题等等。因而《中图法》的修订也被提到议事日程。1996年7月《中图法》第五届编委会在京成立,决定开始对《中图法》第三版进行修订。会议确定了修订的指导思想,制定了修订原则、确立了修订重点、方法与组织。修订指导思想是:

1.《中图法》的修订,应该体现出修订特色,突出重点,保证修订质量,提高其科学性和实用性。

2.《中图法》的修订,必须最大程度地考虑图书馆排架的实际情况,在充分满足其排架要求所需的稳定性前提下,优化《中图法》的检索功能。

3.《中图法》的修订,必须坚持综合性分类法的基本前提,从综合性分类法结构体系的角度,照顾专业单位的需要,处理好集中与分散的关系,保持各学科专业类目深度的基本一致。

4.《中图法》的修订,必须贯彻该法的连续性,在类目类号的编排上,应具有前瞻性和预见性。

修订原则是:

1.保持《中图法》作为列举式分类法的基本属性不变;保持《中图法》的基本部类和基本大类设置以及序列基本不变;保持《中图法》字母-数字混合制的标记符号与层累小数制的标记制度基本不变;在此前提下,进行《中图法》个别大类的类目体系结构的调整与完善,新学科、新事物主题概念的增加与补充。

2.充分利用《中图法》已经采用的分类方法与技术,并不断完善这些技术。如多重列类法,交替类目法,参见类目法、类目复分、仿分法,主类号直接组配法等。

3.《中图法》第四版中应较多地考虑方便用户从第三版改用第四版的措施。如第三版中被移位类目的类号,第四版一般不用,并有改入新类的沿革注释;有关《中图法》第三、四版修订类目的对照表合并进《中图法》第四版;《中图法》第四版使用手册应与《中图法》第四版同时出版等。

4.在出版《中图法》第四版印刷版的同时,推出该分类法的机读版(电子版),并适时修订出版《中图法》的系列版本(简本、期刊版、少儿版、分类主题对照版等)。

5.《中图法》第四版考虑采取排架用类目与目录用类目分离的做法,并采取有效的措施指导用户在分类实践中加以贯彻。

6.在修订《中图法》的同时,修订《中国分类主题词表》。

从1999年3月出版的《中图法》第四版的实际情况来分析,基本遵循了以上的修订指导思想,贯彻了修订原则,新版《中图法》即将在全国图书馆情报机构推广使用。

2 《中图法》发展的主要经验

《中图法》与国内其他分类法相比，编制产生年代较晚，但发展却迅速而稳妥。在其编制与发展过程中，系统地继承了我国各种分类法的编制经验，吸取了国外分类法的编制理论和技术。如继承了《中小型表》、《大型法》创立的"五分法"体系及其所采用的字母与数字相混合的标记符号制，采用了DDC、UDC的十进制、小数制编号制度，借鉴了UDC、ББК等分类法的分面组配技术等，形成了一部既可以组织藏书排架又可以进行分类检索的列举与组配相结合的分类法。

《中图法》在其发展史上获得过殊荣，曾于1985年荣获国家科学技术进步一等奖。曾被美国图书馆协会主持编制的《世界图书馆及情报机构百科全书》收入到"世界九大分类法"词条中。《中图法》的发展历史是在稳定中求变、不断完善的历史，体现在以下几个方面：

2.1 《中图法》各种版本不断出现与发展完善

为了适应不同类型图书馆、不同类型和不同专业文献分类的需要，《中图法》编委会下设了相应的分编委会和专业分类法编辑组，编辑出版了《中图法》的不同版本和有关的辅助工具书。

《中图法》是供大型综合性图书馆及情报机构编制分类检索工具、组织分类排架的规范工具，已经出版了四个版次。专供藏书较少的中小型图书馆《中图法》简本做到了与基础本同步修订、出版，第四版正编辑修订中，即将出版。《中图法》简本也可供计算机网上资源分类使用。

为适应中外文期刊分类排架和分类检索的需要，1987年开始出现了《中图法》的系列版本——《中图法·期刊分类表》(简称《期刊表》)。

为适应中小学与儿童图书馆类分藏书的需要而编制的《中图法·儿童图书馆、中小学图书馆版》(简称《少图版》)也于1991年、1998年分别出版了第一版和第二版。为了满足各个专业图书馆文献细分的需要，出现了与《中图法》体系、结构、编制原则标记符号基本保持一致的系列专业分类法。如1993年出版的《中图法·教育专业分类法》，1995年出版的《中图法·公安科学文献分类表》，即将出版的《中图法·医学专业分类法》，《中图法·农业专业分类表》等等。

为实现分类主题一体化标引和检索，为机助标引、自动标引提供条件，降低标引难度，提高检索效率和标引工作效率，作为一种新型的分类检索语言和主题检索语言兼容互换的工具，同时作为《中图法》的系列工具出版了，它就是《中国分类主题词表》。

2.2 在保证图书排架稳定性的前提下，不断加强《中图法》检索功能

《中图法》从第一版出版起，在我国文献分类实践中就同时用作组织分类排架和组织分类目录的规范工具。但是分类排架和分类检索对分类法有不同的要求，前者要求类号唯一性、稳定性和简单性，而后者需要提供多角度、多方面、多层次的类号检索，而且类号需要有灵活组配性。因此，如何在保证图书排架稳定性的前提下，加强其检索功能，一直是《中图法》编者们探讨的问题。在《中图法》各版的修订中都特别强调分类法满足排架的稳定性要求，在修订第三版过程中，提出了分类排架用类目与分类检索用类目分离的设想，但由于种种原因，最后未能实施。遗留了一个待解决的课题。加强《中图法》检索功能的措施主要有：

1.扩大多重列类法的应用。即同时采用两个或两个以上分类标准对某个大类或某个类目进行划分，形成两个类目体系或多个类列，增加了类目不同属性主题的揭示，有助于提高文献更多的检索入口。如U448各种桥梁、J41各种摄影艺术等。

2.扩大组配技术的应用。采取增加通用复分表与专类复分表的数量及其表中类目的数量、增加主类号之间的组配、增加类目之间的仿分等,目的在于提高分类检索的专指性和网络性以及灵活性。如第四版增设了"世界种族、民族表",增加了12个专类复分表。增加了形式区分符号("""、()）的使用范围。

3.扩大交替类目和多分类体系的应用。设置交替类目并扩大到某一分类体系的交替,形成局部的交替分类体系,如TE12、TE13等类。这样一方面可满足专业文献分类的需要,另一方面也可满足《中图法》用于加强分类检索功能的需要。

4.加强资料分类检索的功能。为了适合类分资料的需要,《中图法》第四版把用于资料分类的类目与用于图书分类的类目共存于一表,用符号和注释来分离。进一步加强了《中图法》的文献检索功能。

2.3 不断加强类目的语言描述与注释行文以及标记符号配置的规范性

类目语言描述的规范性包括类目含义划分的逻辑性和类名、类目注释中自然语言描述的简洁性、准确性和统一性。标记符号配置的规范性主要包括标记符号的逻辑性、规律性、唯一性和特指性。

《中图法》从第二版修订就重视类目语言描述与注释行文的规范性问题。随着机读编目的实践、自动分类技术的研究以及应用计算机技术维护管理《中图法》的要求,规范研究得以更加重视。《中图法》第四版规范了不少类目名称,如将"B哲学"改名为"B哲学、宗教"、"F0政治经济学"改名为"F0经济学"、"TK动力工程"改名为"TK能源与动力工程"等,以准确地反映类目所概括的内容范围;规范了注释用语和注释引用次序,明确并统一了注释用语的行文,规定了注释的引用先后次序,如将"入此"注释与"入他类"注释分段行文;含义注释位居最前、沿革注释位居最后等。

对标记符号的规范方面,不仅制定了标记符号贯彻连续性、前瞻性和预见性的原则,而且加强了标记符号的逻辑性、规律性、唯一性和特指性的要求。如相同属性的类目对应配号,第四版消除前几版中,同一主题资料分类和图书分类时,除类号详略外的其他差异问题等。

2.4 不断加强科学管理,方便用户使用

《中图法》第三版出版后,面临新旧分类法的更替和使用新分类法的具体问题,《中图法》编委会编制出版了《〈中图法〉二、三版修订类目对照表》;自1993年开始定期编制出版《〈中图法〉信息简报》,以方便对《中图法》连续性修订和加强对《中图法》用户的指导。

从《中图法》第四版可以看出,进一步加强了方便用户的措施,如在类表中采取增加"沿革注释"和设置"停用类目"的措施,前者使用户了解类目修订的来龙去脉、掌握修订情况,便于改编已入藏的文献;后者尽量避免两版分类法类号的冲突,为用户改编文献提供了时间上的缓冲。同时继续保持与用户建立联系卡,以便科学维护管理《中图法》。

3 《中图法》发展的展望

分类法是图书馆与情报机构组织文献信息历史悠久的、重要有效的工具。现代意义上的分类法也有120多年的历史。随着众多图书馆处理文献资料的计算机化以及纸制文献载体的数字化,《中图法》面临如何适应计算机文献编目以及应用于网络环境下的在线编目问题。《中图法》在跨入21世纪的当前正在着手如下几方面的课题:

3.1 加强《中图法》发展的理论研究

理论指导实践，实践促进理论的研究与发展。《中图法》在其20多年的发展过程中，积累了一定的经验，应该将已取得的实践经验上升为理论，指导《中图法》今后的发展。

有必要成立专门的、类似FID/CR(国际文献联合会/分类法研究委员会)、CRG(分类法研究小组)等研究机构，或强化《中图法》编委会的研究职责，形成专门的研究队伍，开展与《中图法》有关的专题研究。诸如分面分类理论与技术、分类主题一体化理论与技术、分类法在线应用理论与技术、分类法的计算机化与电子化、分类法的机读数据格式、利用《中图法》进行机助标引、自动标引等需要进一步研究。

3.2 尽快研制《中图法》的电子版

研制《中图法》的电子版是当务之急，也是《中图法》编委会目前正在进行的工作。

《中图法》的电子版有着纸制版许多不可相比的优越性：电子版分类法中可方便、迅速地用分类号或主题词(或关键词)查找；可进行全部查询词或分类号列表；可实现全文检索、组合检索、限制性检索，扩检、缩检与相关类号跳转极其便捷；多屏幕显示，按任意要求的等级显示或其他显示；电子版的分类法也比纸制版的分类法价格便宜。总之，分类法的电子版是发展方向。

国外分类法电子版已得到广泛应用的有OCLC研制的“Dewey for Windows”(一部电子版的DDC21版)与“Classification Plus”(实现了把国会分类法与国会标题表共存与一张光盘，用于在线检索)。随着《中图法》的纸制版出版，《中图法》的电子版已着手进行，设想先开发出单机版与局域网版，待条件成熟，可将《中图法》放至广域网供用户使用。电子版的出现，必将带来图书馆信息机构处理文献上的深刻变革。

3.3 着手研制《中图法》的机读格式，制作《中图法》的机读数据

从计算机编目工作角度来看，主题数据与分类数据是书目数据中的规范数据。根据IFLA UBCIM 1987年出版的UNIMARC/Authorities Format，国家图书馆已于1997年研制出中国机读规范格式(其中包括主题数据的规范格式)，目前已作为文化部的部颁标准申报，国家图书馆根据该格式与《主题规范数据款目著录规则》，将《中国分类主题词表》中的全部主题词转换成了符合ISO2708格式的标准机读数据，可作为通用数据供交换和使用。分类法规范数据方面，美国早在1990年就研制出了分类法数据的USMARC格式，并在1995年作了修改。IFLA起步较晚，于1993年成立了工作组在研制分类法数据的UNIMARC格式，目前已有一份正式的研究报告，名为“分类法数据的格式要求”，报告中提出了三种可供选择的方案：采用现有的分类法数据MARC格式，扩展UNIMARC/Authorities格式，以适应分类法数据需求，开发分类法数据的UNIMARC格式，预计在今年内会有最后的结果。

由于IFLA尚未确定分类法数据的标准格式，因此，目前《中图法》编委会正在研制的《中图法》机读格式还只能是临时性的，不过，这种临时性的格式参照了IFLA的主题数据规范格式和美国的分类法数据USMARC格式，可望作为我国的分类法机读格式的基础。

《中图法》数据的机读格式与我国的书目数据格式、规范数据(名称与主题)格式等兼容。遵循“文献目录信息交换用磁带格式”标准。数据元素主要分为0—标识块；1—编码信息块；2—分类号块；3—类目注释块；4—无效类号根查块；5—有效类号根查块；6—其他分类号块；7—主题词块；8—来源信息块；9—国内使用块。通过指示符来描述类号的来源和类号的性质；通过子字段来描述类

号、类名、类级；通过不同的字段来描述类目的注释类型、类与类之间的关系、类目与主题词之间的关系、分类法之间的关系等。

《中图法》数据的机读格式可作为《中图法》电子版的条件，也是建立计算机维护管理系统的基础，二者之间可相互利用，共同促进《中图法》的发展。

3.4 建立《中图法》与《中国分类主题词表》的计算机综合维护管理系统

从《中图法》更新版本以及日常维护管理来看，目前由于缺乏计算机管理系统，因而无法完成消除类目概念重列、类号类级排版错误，不准确的类目参照关系等现象，使《中图法》的质量受到一定程度的影响；从《中图法》各系列版本同步更新发展来看，由于缺乏计算机管理系统，很难达到同步一致的修订。

《中图法》与《中国分类主题词表》的计算机综合维护管理系统将包括数据录入、逻辑查错及质量校验、数据维护与更新、数据统计、编辑排版、检索与查询、电子分类表生成等多项功能。

通过建立上述管理系统，可以使《中图法》以及《中国分类主题词表》及时维护与更新，为文献的自动分类标引与自动主题标引创造条件、打好基础。

社会在发展，人类进入了一个数字化生存的时代，分类法必须适应网络环境下的文献信息处理与组织，也许一部分类法的内容框架将发生改变，或根本性的改变，但对于组织文献、知识与信息来说，分类法的原理永远也不会过时。《中图法》在目前和将来还需进一步发展。

参考文献：

[1]张琪玉.情报检索语言走向自动化之路与《中图法》发展目标.北京图书馆馆刊，1996(4)
[2]卜书庆.评《中图法》.北京图书馆馆刊，1993(2)
[3]汪东波.《中图法》第三版存在的问题及第四版的建议.北京图书馆馆刊，1996(3)

面向21世纪《中图法》及其标引技术改造的研究

华东理工大学石油化工学院图书馆　陈树年

[摘要]本文通过对体系分类法适应计算机检索能力的分析，针对《中图法》改造的现实性、可行性进行了研究，提出了对《中图法》类目结构和标引规则进行改造的基本方法，分析了改造效果及存在的问题。

[关键词]《中图法》　类目结构　标引规则　改造

1　体系分类法对计算机检索适应能力的分析

现代分类法自杜威创制《杜威十进分类法》已经历了120多年，这期间有体系分类法、分面组配分类法、体系－组配式分类法三种基本类型并存，但体系分类法始终占据着主导地位。1933年阮冈纳赞的《冒号分类法》问世，尽管分面组配理论引起巨大的反响，但这种分类法似乎过于复杂和神秘，面、亚面、点、巡、层、相、轮等生涩的概念加上令人目眩的组配公式，不仅将主题分析过于复杂化，增加标引的难度，更重要的是用户极难掌握，其实用性大大降低，这是它不能广泛被接受的主要原因。1905年诞生的《国际十进分类法》，在体系分类法的基础上大量运用组配技术，成为体系—组配式分类法的代表。由于UDC的辅助标记繁多，组配面成的类号可能相当长，显然不太适用于普通图书的分类，更不用说图书的排架了。长长的UDC类号中可以嵌入各种主题因素，虽可用于机检，但不论文献标引、文献检索，还是计算机处理，都是很复杂的。

建国后我国的分类法曾努力向CC、UDC学习，尽管在编制新型分类法方面做过各种尝试，但始终保持着以体系分类法为主的格局，看来这和国情与习惯是有关的。在计算机技术普通应用于图书情报工作领域的今天，传统的体系分类法在适应计算机检索方面暴露出很多不足。另外新编分类法和对现行分类法进行重大改造都是不现实的，那么如何在目前通行的分类法的基础上，经过不"伤筋动骨"的改进(包括标引规则)，使之满足计算机检索的需要，就成为检索语言研究的一个重要而现实的课题。下面对这一问题作初步的探讨。

1.1　体系分类法的基本原理及形成的基本功能

体系分类法编制的基本原理可以概括为三方面，一是聚类的原理，即文献按什么属性区分和类聚，由此产生相应的基本检索途径，"学科分类和知识分类"是体系分类法聚类的基本依据，从文献的内容对文献加以区分这个角度来说，它与以事物为中心聚类的主题法形成明显的分工；二是类目组织的原理，即怎样建构分类法的类目结构，包括类目划分中分类标准引用次序的使用、类目关系的显示、类目排列方法等，类目组织既是实现科学分类和知识分类聚类的手段，又是确定各个门类检索途径的手段，"逻辑划分原则"是体系分类法组织类目的基本方法；三是标记原理，即如何为类目配置固定的代号，做为标记类目次序、组织分类检索工具的依据，"先组式类号"是体系分类法类

目标识的基本方法。

以科学分类和知识分类聚类,使文献按其内容的学科属性、知识属性加以区分和集中,形成了系统区分的体系,提供了分类法按学科分类、知识分类、专业分类查找文献的基本途径。逻辑划分规则使类目的组织自上而下,以列举的方式逐层划分、层层隶属,形成一个严密的等级结构,充分揭示类目之间的逻辑关系,使学科分类和知识分类以一种具体形式展示出来,形成一幅知识地图。在这个类目等级结构中,每个类目都具有表达知识概念的唯一性,它们的外延一般不相互交叉。先组类号标记方法,使每一个类目都有唯一的标记,并以线性的方式排列先后次序。体系分类法的这些基本功能,使它可以很好地对文献内容从知识分类角度加以描述、对文献(包括知识单元)和分类款目进行线性组织,较好地满足手工检索的需要。

1.2 逻辑划分规则与等级列举对计算机检索的制约

逻辑划分是对一个宽泛概念逐次增加内涵、限制外延,使概念越来越专指的过程。增加内涵,就是使用概念所具有的属性做为分类标准对其进行分类,这个过程中是连续的,并且一次划分只能使用一个分类标准,同一次划分出来的子概念之间是平行的。体系分类法采用等级列举的方法表达知识体系中的内涵深浅不同的众多概念,一个概念由于在连续划分中不断增加新的内涵,它的涵义也越来越复杂,类目(概念)中的主题因素越来越多。每个主题因素都是用一定的语词来表达的(尽管体系分类法采用下位类类名承接上位类内涵的方法简化类名),那么随着分类层次的增加,表述复杂概念的类目其语词结构也复杂。不管一个类目包含多少主题因素,它们一经形成一个专指的概念,便从此构成一个凝固的整体,也就是先组式的类名,任何时候都不能再分解开来使用。例如"U464.337.2 点火装置"(以《中图法》为例,下同),它实际的涵义和包含的主题因素是"公路运输—汽车工程—汽车发动机—燃气轮机—部件—起动系统—点火装置"。这是一个完整而凝固的主题因素链,并且只能用来表达某个特定的主题,除了按规定可与某个类目(独立的主题概念)进行组配之外,各个类目中的各个主题因素是不能加以分解和用于组配的。换言之,运用逻辑划分规则和等级列举方式编列的类表,分析与综合的功能是十分有限的。它所适应的编目方式、检索方式、编目和检索设备,都是手工式的。计算机技术的应用使文献编目和文献检索发生了重大变革,我们的观念也必须随之改变。捷克学者托曼(Toman,I.)曾深刻地指出:"对字顺主题和分类系统的研究表明,任何一种系统只不过是对同一组织原则进行另一种组合而已","当这些系统用于电子计算机时,所有的差别都消失了"。对于计算机检索来说,它除了具有运算速度块、存储量大外,还有几个重要的特点:

第一,它不像图书排架、目录排列那样是线性的,在机读目录中,一条记录著录有文献的各种数据和多种检索标识,并分别编成索引文档,它们之间是平行的,分类目录、主题目录、题名目录、著者目录的界线已经消失。这种文档结构,特别适合文献各要素之间的逻辑运算,也就是组配检索,这就需要检索语言能提供数量众多的、能参与逻辑运算的主题要素,而体系分类法凝固的主题因素链制约了这种功能。

第二,计算机检索是面向用户的,如果说早期的计算机检索多数用户需通过馆员这个中介才能进行,那么现在多数用户则直接上机检索(甚至在自己的家中),面向21世纪的检索语言必须充分考虑和满足这种变化了的需求。体系分类法的文献排架功能主要是面向馆员的,文献检索功能则是面向用户的,姑且不说它复杂的编号、组合规则用户无法掌握,单是它体系的严密和复杂性、类名涵义的隐含性、细小主题的隐蔽性就足以使用户望而生畏。一个图书馆专业毕业的本科生,如不经过一段扎实的实践过程尚不能十分熟悉他所使用的分类法,更不用说一般的用户了。实际上用户

在使用手工分类目录检索时带有相当大的盲目性和随意性，其检索效率也就可想而知了。对于计算机检索用户来说，宽泛的分类浏览检索往往使用较概括的类号进行，较专指的分类检索则希望使用已知类号进行组配检索；用户还有一种检索需求，就是根据他已掌握的或需要的文献要素进行组配检索（不管是分类的、主题的、著者的、题名的）。现有的体系分类法及其相关的标引规则，都对这种检索需求产生制约作用。

体系分类法先组式的标识系统是与它先组式类目体系相对应的，一个概念的每一级逻辑划分，都需要在原有类号的尾部上增加一位或两位符号来表达。同类目主题因素的结合一样，类号的结合也是次序固定的、凝固的，形成一个个表达特定主题概念链的符号链。以上例的“U464.337.2 点火装置”为例：

公路运输—汽车工程—汽车发动机—燃气轮机—部件—起动系统—点火装置

U4 ——— 6 ——— 4 ——— 3 ——— 3 ——— 7 ——— 2

这种对列举类目的标记方式，使每一个列举的类目都有一个固定的类号，类号逐级从尾部展开，类号的长度基本与类目的细分级别一一对应：

U4　公路运输

U46　汽车工程

U464　汽车发动机

U464.3　燃气轮机

U464.33　部件

U464.337　起动系统

U464.337.2　点火装置

由于每个类目都配有独立的分类号，因此都可以做为独立的检索入口，如果需要的话，也可以进行主类号之间的组配标引和组配检索。除了因划分层次较多，用户不容易掌握之外，是能满足计算机检索需要的。与计算机标引和计算机检索不相适应之处，出在体系分类法的组配技术的运用上。

1.3　体系分类法组配技术的局限性

组配，是将两个或多个简单的要素，按一定的规则组合成一个复杂要素的方法。在叙词语言中是指将两个或多个已有简单概念合成一个新的复杂概念；在分类语言中是一种类号（实质是类目概念）的合成技术，按一定的规则或使用特定的组配符号，利用已有的类号（包括复分表的子目），合成一个新的更专指的类号。组配技术既用于文献标引，也用于文献检索。

体系分类法为提高其灵活性，在传统的等级列举编类的基础上，也在不断引进组配技术的应用，并呈现扩大的趋势。在体系分类法中运用组配技术基本有四种类型：

(1)主类号仿照已列出的有关类目细分。

例如：“法国的科研生产联合体”可标引为 F279.565.442，其中“565 ”（法国）是依世界地区表分的，“4”（各种企业经济）是仿 F279.2 中国企业经济分的，“42”（科研生产联合体）是仿 F276 各种企业经济分的。

(2)主类号使用专类复分表的子目进行细分。

例如：“跳高的生物力学”可标引为 G823.114.6，其中“14”（体育运动基础科学）是专类复分表的子目，“6”（生物力学）是仿 G804 体育基础科学分的。

(3)主类号使用通用复分表的子目进行细分。

例如："美国对晚期肺癌治疗研究的最新进展"可标引为 R734.205(712)=5<147>，其中"05"（治疗）是依专类复分表分的，"(712)"美国、"=5"现代、"<147>"晚期是使用特定的组配符号与世界地区复分表、国际时代表通用时间地点复分表的主题因素进行组配的。

（4）主类号之间直接进行组配。

例如："图书馆经济学"可标引为 F08:G250。

第一、二两种组配形式，主要的作用是提高类目的细分能力，也就是在不增加类表篇幅的条件下提高分类法描述未列举主题的能力。

第三种组配形式，除了具有第一、二种组配形式的功能外，还具有潜在的将用特定符号连接的主题因素分解出来的功能。

前三种组配形式由于提高了标引的专指度，从理论上讲可提高检准率。但实际上体系分类法结构的复杂性和组号的复杂性，使一般用户很难掌握，检准率的提高就会大打折扣。上例类号虽是组配而成的，但是先组性的，一旦合成后就成为一个凝固的整体，其中所包含的各个主题因素便不能离析出来（特别是第一、二种方式合成的类号），因此不能作为独立的检索入口使用，多途径检索也就无从谈起。所以它们的真正功能在于缩小类表篇幅和提高描述新主题的能力，而对提高检索效率没有多大的影响。

第四种组配方式是比较灵活的，除可用于类目细分之外还可合成新的主题概念，但这种组配在手工编目、手工检索和现有类目结构条件下，不论从标引还是从检索角度来说其作用都是十分有限的，原因是：

一是在手工编目条件下，分类号的长度受到限制，不可能用很多的主类号组配成长长的字符串，例如"用于化纤织物染整自动控制的模式识别装置"这一主题，标引的结果是 TS190.645:TP273:TP394.1，这样对组织文献和组织分类款目都不方便。

二是体系分类法编制受逻辑划分规则和划分层次的制约，文献主题中所有具有检索意义的主题因素（属性），不会都作为分类标准使用，也就是没有描述这方面主题因素的类目，那么使用主类号组配某些新主题概念成了"无米之炊"。此外，为保证现有类目体系不受组配编号的"冲击"，主类号直接组配法从标引规则上也严格地进行了控制。

三是在手工检索工具中，分类号的轮排不但加大了检索设备的体积，也大大增加了排片的工作量，很难加以推广，而分类号如不进行轮排，提供的检索入口还只是一个，主类号直接组配标引的检索意义也就荡然无存了。

四是在手工检索条件下，无法实现后组配式的逻辑检索，用主类号逻辑组配起来的类号只能作为先组类号使用，即使进行轮排，也只相当于附加分类号的作用。不可能利用众多组配类号中的众多主题因素，根据检索需要任意进行组配，产生所需的文献集合输出。

从以上分析不难看出，尽管现代体系分类法也具有很多的组配功能，但是由于受类目结构和标引方式的限制，其适应计算机检索的性能还是很差的。

1.4 多重列类及其现有标引方式对计算机检索制约

等级体系分类法是我国各种分类法编制的基本模式，这种分类法用建立等级结构的方式组织和显示人类的全部知识。为揭示各知识领域及各学科门类的从属关系、并列关系和它们之间亲疏远近，首先确定一个知识分类大纲即基本大类，然后以基本大类为起点，逐级进行概念划分。在对一个较大的概念进行划分时，遵循逻辑划分的规则，即每一层次的划分只能选择使用一个分类标准（事物的某种属性），使划分出来的子目（子概念）增加相同的内涵，从而限制了其外延。通过逻辑划

分形成的子目之间没有外延交叉，这对类分文献是十分实用的。然而事物的属性是多方面的，当某事物的几种属性都具有检索意义时，应首先使用哪种属性进行划分呢？于是就产生了"分类标准的引用次序"，即人为地确定类目划分时所使用分类标准的先后。虽然某事物所具有的几种属性都作为分类标准被使用，但文献只能按最先使用的分类标准集中，第二划分层次的分类标准仅能在第一层次划分形成的若干区间分别集中有关文献，越后使用的分类标准按该属性集中文献的能力越弱，使用该属性作为检索入口检全文献越难。

为了使事物所具有的多种属性在一个相对较高的划分层次（通常选择某类事物的类称）中同时作为分类标准使用，避免在多层次划分中被先使用的分类标准"层层切割"，减少聚类的能力，于是等级体系分类法采用了变通的方法，允许部分类目在同一划分层次同时使用几个分类标准进行划分，这就是所谓的"多重列类"。这样在一个上位类下就形成几组并列的、外延有交叉的同位类。例如：

U 448　各种桥梁
448.1　各种用途桥梁
448.11　人行桥
448.2　各种结构桥梁
448.21　梁式桥
……………………

"桥梁"在这里是事物的类称，以此为多重划分的起点，就可以把各种属性或特征的桥梁同时平行地列举出来。其作用在于：如果分别从不同的属性论述桥梁，都可以选择相应的类号进行标引，用户也能以不同的属性作为检索入口检索文献。但如果文献主题同时涉及事物的几种属性，由于同位类存在外延交叉，就有几个类号可同时用于标引，为保证同主题文献标引的一致性，就人为规定采用"最前编号法"或"最后编号法"。这时，多重列类的优越性就无法体现。

采用分面分析产生的一组组性质纯粹、内涵相对较浅的类目（与叙词接近），主要作用是用于对文献进行组配标引，并为机检的后组检索提供条件。因此，在体系分类法中多重列类法不与组配技术相结合时，其作用是十分有限的，主要是在一定程度上提高对文献主题的描述能力，而对检索效率影响也是不大的。

1.5　体系分类法"集中与分散"矛盾的分析

众所周知，"集中与分散"的矛盾是体系分类法的基本矛盾之一，产生"集中与分散"矛盾的根本原因，是体系分类法编制技术所决定的。体系分类法是等级列举式分类法，编制的基本技术是对概念进行层层逻辑划分，形成层层隶属的概念等级体系。由于客观事物都具有多种属性，以某种属性为分类标准，便可以与具有相同属性的事物构成一类，即有"多向成类"（或"多向成族"）性。体系分类法类目划分时，一次只能使用一个分类标准，即选择事物的某一属性与相关事物成类，而且，在一个学科或专业范围内一旦与某一事物组成一类后，便不能再以其他属性与别的事物成类。事物本身"多向成类"的客观性，在体系分类法中强制成"单向成类"的人为性，由此产生了"集中与分散"的基本矛盾。其基本表现是：当事物按某种属性集中成类后，对于按其他属性组成的类就分散了；当事物按专论集中时，按总论就分散了，反之亦然；当事物按属种关系集中时，按其"方面"就分散了；当事物具有的多种属性都被作为分类标准时，只有按首先被引用的分类标准（属性）集中，而对于后被引用的分类标准（属性）还是不同程度的分散了。也就是说，除了在较大的范围内按学科和专业可以基本检全文献外，从某种属性或事物出发无法检全文献。

一般认为，体系分类法“鸟瞰全貌”、“触类旁通”的优点和产生“集中与分散”的矛盾都是出于同一原因，即对概念进行层层划分、层层隶属的结果。因此，从根本上改变这种状况是不可能的。

文献标引的目的有两个，一是组织文献，二是检索文献。对于组织文献来说，体系分类法是理想的工具，除了按学科集中文献外还能满足文献单线排列的需要，这是其他分类法所不及的。但对于检索文献来说（这是分类法的主要功能），由于它不能按不同的属性分别集中文献，难以提供多途径检索，是其重要的不足。在手工编目和检索条件下，要从根本上改善是不可能的。

在手工编目和检索的条件下，不能解决按不同的属性分别集中文献、实现多途径检索的基本原因是：①列举类号是一个严格且凝固整体，其中的各个主题因素除排列在最前的均不能作为独立的检索点；②标引阶段利用仿分、复分组配的类号，符号组合顺序是固定的，组配时所使用的主题标识也是截取某类号的局部或使用无独立检索功能的复分号，都不能作为独立的检索点使用；③使用主类号直接进行组配，一是不论从文献排架还是款目排列来说都不允许类号过长，限制组配的级别；二是进行全面轮排是不可能的，而不进行轮排，则不能按不同属性集中文献；④使用“互见分类”也因涉及排卡工作量而不能广泛采用；⑤《中图法》、《科图法》、《人大法》等体系分类法由于主要立足于手工编目和检索，在类目编列和标引规则上也不能很好地满足这种需要。

2　对《中图法》改造的现实性、可行性及基本思路

2.1　对《中图法》进行改造的现实性和可行性

改造旧分类法使之适应新的需要，还是站在新的起点和高度编制全新的分类法，这是情报检索语言界经常提出和讨论的问题。谁都知道，成功地改造一部旧分类法要比编制一部新分类法的难度高得多，尽管新编的分类法在体系结构、功能等方面可能都优于旧的分类法，但人们通常还是要选择费时、费力修旧，而不选择省时、省力造新的途径，这可以说是情报检索语言领域中的一个特殊现象，归根到底这是图书情报工作特殊性所产生的。

《中图法》（包括它的系列版本）自 1975 年诞生以来，很快被我国图书情报界所接受，应用的领域越来越广，除绝大多数图书情报部门使用《中图法》组织藏书排架和分类检索工具以外，各种书目、索引、文摘，各种文献数据库，也多以《中图法》进行分类。这样，不但我国各类图书情报部门的工作人员对《中图法》的分类体系是熟悉的，我国广大读者对《中图法》的分类体系也是熟悉的，这是一个十分重要的基础。如果编制一部新分类法代替《中图法》，不但涉及数以亿计的图书馆藏书的改编问题，也涉及改变读者和图书情报部门工作人员使用分类法的习惯，显然这是十分困难的。因此，从我国的具体情况来看，对《中图法》不断进行改造、完善，使之满足新的标引需求，是一条十分现实的路子。杜威十进分类法、国际十进分类法、美国国会图书馆分类法等这些历史悠久的分类法，尽管都存在这样或那样的不足，但它们都在不断的改造和修订中顽强地保持着自己的生命力，还没有一部新编分类法能取代它们，都说明图书情报系统使用检索语言有着巨大惯性。

对《中图法》进行改造以更好地适应计算机检索，不仅具有现实意义，更重要的是具有很好的可行性。做为既能满足图书馆组织藏书，又能满足计算机检索需要的分类法，应具备以下几个基本的条件：

(1)有一个较理想的知识分类体系，能科学地组织人类全部知识的载体——文献。

(2)有一个能系统组织和揭示各学科门类知识的等级结构。

(3)有一个既满足知识体系展示和扩充需要，又兼顾组织文献需要的标识系统。

(4)有一个能充分描述各学科门类知识,先组度不宜过高、规范的术语系统。

《中图法》是建国后编制较成功的一部综合性体系分类法,上述这四个基本条件是大体具备的,它在组织藏书、编制手工检索工具方面的功能较为完备;在满足计算机检索方面,它的标识系统和术语系统却存在一些问题,但改造起来并不困难。

2.2 改造《中图法》,使之适应计算机检索的基本思路

(1)我国的图书情报工作长期以来都是传统手工条件下进行的,由此不但形成一套成型的工作模式,也逐步形成一种看待各工作环节及其关系的僵化的思维方式。这种传统的思维方式,对我们今天广泛而深入的应用计算机技术、通信技术处理图书情报业务,往往会成为一种无形的桎梏,束缚我们大胆地思考问题。要摆脱长期手工标引和手工检索形成的对分类法改造及编制的束缚,充分借助于计算机强大的存储能力和运算能力,完成手工检索所不能进行的工作。传统的分类法应向以机读分类法和联机分类法为主的形式过渡,印刷本应作为机读版的附属形式。机读分类法修订方便,动态性强,能紧跟世界科学技术高速发展的步伐。

(2)改造后的分类法仍具有组织文献和检索文献的双重功能,但检索功能是主要的。保持用于检索的分类号中的一个与文献排架号相同,使分类法的改造不影响已入藏图书的排架。适当降低排架分类号的细分程度。

(3)保持体系分类法固有的系统组织知识和检索知识的功能,同时大大提高其多途径检索的功能,为用户通过浏览缕清情报需求或产生新的情报需求提供条件。

(4)在计算机检索中,用户越来越减少对专业人员中介服务的依赖,而直接介入检索。分类法应把检索人员和标引人员作为同等重要的用户,因此分类法的改造应向着结构简明、透明度高、使用方法简单的方面发展。

(5)为满足信息社会“知识爆炸”对检索语言标引能力的日益提高的要求,体系分类法应提高对文献主题的描述能力,包括描述的专指度、描述新主题的能力。

(6)组配检索是计算机检索中重要而极易实现的功能,分类法的标识系统只有做到每个标记符号都具有独立的标引和检索意义,才能充分实现在计算机系统中的组配检索。

(7)网络化是信息传递、交流、获取的必由之路,世界范围的信息资源共享程度将越来越高,这要求分类法在改造中把兼容性作为一个重要因素来考虑。

3 对《中图法》进行改造的要点

一部分类法的功能,不仅取决于类表编制的方式和质量,也取决于与类表相互配合的标引技术。对分类法的改造必须将分类表与其对应标引技术结合起来进行。

3.1 分类法的体系问题

系统检索功能是分类检索语言基本功能,也是与主题检索语言的最重要的分工,基于机检的体系分类法也必须充分保持这种功能。基本序列和大类结构的理论,一百多年来这个问题讨论的不少,各个分类法都有一套自己认为是正确的理论,虽都不统一,但也能完成建立类目体系的任务,可见这个问题对检索效率的影响是不大的。分类法是一种工具,其实用性是第一位的,它的基本类目结构,应与当代的学科结构大体一致,历史上曾有重要影响并有大量文献的学科或主题,也应以相应的形式保留在分类体系内。科学界所公认的分类、人们所习惯的分类、国际及国家所颁布的学科

分类、高等学校稳定的专业设置，应是基本大类设置的主要依据，至于基本大类的排列顺序，有一个基本合理的次序即可，也不必为"历史"在前或在后、"经济"在"政治"之前还是在其后进行无结论的探讨。适于机检的分类法的基本大类设置要完整，并与文献的现状和发展趋势相适应，根据科学的发展有控制地增加基本大类是必要的。类目的展开要层次分明、清晰，同一级的划分尽可能充分，只有保证良好的结构性才能满足系统检索的需要。

3.2 提高分类法描述文献主题的能力

传统的体系分类法编制，因过于拘泥逻辑划分的规则，又受制于划分的层次，因此某一事物的多种属性只能做到有限的揭示，当一事物的某种属性未作为划分标准，就不能从这个角度描述文献，那么也就不能从这个角度去检索文献。提高对主题描述能力的基本方法是采用分面分析的技术，对一个事物的同一度划分分别采用多个标准进行，也就是我们通常所说的"多重列类"。

广泛使用多重列类法，最重要的优点是使更多的主题概念进入分类体系，不仅大大提高体系分类法描述主题的能力，而且，为类号直接组配创造条件，形成了在机检条件下按任意属性检索文献功能的基础。因此，在不影响其系统检索功能的基础上，尽可能多地列举能独立描述主题单一属性的类目。多重列类技术如大量在较高的级位上使用，将影响体系分类法的系统性，因此以在三级以后类目的编列中使用为宜。

3.3 取消"专类复分表"，代之以"一般性问题"或总论性类目

专类复分表的作用是作为专论性子目共性问题复分的依据，由此带来的最大弊端是将专类复分表中的各主题要素转换为号码后，"焊接"在主类号的尾部或中间，完全失去了检索的价值。由于《中图法》编列的专类复分表绝大多数是该类总论性类目简化形式，并且配号一致，取消专类复分表后，不但可以兼容原有按专类复分表复分的类号，还可以提供文献标引专指度，更重要的是增加了新的检索入口。例如：

专类复分表	主表类目
01 演化与发展	Q931 微生物的演化
02 细胞学	Q932 微生物细胞学
03 遗传学	Q933 微生物遗传学
04 形态学	Q934 微生物形态学
…………	………………
09 分类学	Q939 微生物分类学

少数专类复分表没有对应的总论性类目，也很容易补编。

3.4 除了直接描述事物、问题的概念外，提供足够的限定主题的一般概念，可提高对主题概念限定的能力，从而提高检准率

其措施是设置多种类型的通用复分表，并详细列举子目。《中图法》第四版虽新增设了"世界种族与民族区分表"，并对各通用复分表不同程度地进行了扩充，但这还不够，应当在通用复分表的改造上下更大的功夫。比如增编有关人物、材料、语言等通用复分表。

3.5 建立完善的参照系统和详尽的注释系统

参照系统和注释系统是揭示分类法类目之间关系、类目含义，也是用户掌握和使用分类法进行标引和检索的重要手段。完善的参照系统可向用户揭示与某一类目相关的一切的类目，帮助用户选择最合适的类目，避免遗漏。详尽的注释系统有助于用户确切了解类目的主题范围，特别是类名字面所没有明显揭示的主题。机读分类表可使用户随时从参照指引的主题或类号、注释中指引的主题或类号，快速切换到相关的类目，进行甄别比较，决定取舍。

4 对《中图法》标引规则改造要点

不同的标引技术方法，对发挥同一部分类法在计算机条件的检索性能影响甚大。根据我们对《中图法》现行标引方法对计算机检索不适应性的分析，可以把改造的思路概括为：

①尽量为用户提供新的检索入口，增强多途径检索的性能；

②提高文献分类标引深度和分类标引专指度；

③尽可能为发挥计算机检索系统的后组配检索功能提供条件；

④通过标引技术的改进，缓解或局部消除体系分类法的"集中与分散"的矛盾；

⑤有条件地把类目的仿分、复分改为类目组配标引；

⑥将部分凝固的先组类号(即类目中固定的主题要素组合)加以分拆，以形成独立的检索点；

⑦利用通用复分号前置区分符号的区分功能，使通用复分要素成为能与主类号实现后组配检索的独立类号。

4.1 改变《中图法》的组配组号方式，以适应机检后组配检索的需要

《中图法》的分类标记组合中，也大量使用组配方法合成类号，但实际上这种组配技术的应用主要局限于扩展类目体系，而不是用于检索。这种方法在机检条件下是不可取的，一是这种组配复分十分复杂，不但用户就是标引人员也不易掌握，由此构成的长长的类号串也不便于藏书组织；二是由此产生的类号虽然包含多种主题因素，但这种类号一经组成，便成为一个凝固的整体，各组成部分不再能分拆离析，独立地表达某种主题的属性，也不能作为一个用于组配检索的主题因素加以使用，这种类号组配方法是它不适应计算机检索的致命弱点。

改编组配组号方式的目标是：把《中图法》运用各种组配方法合成的包含多种主题要素的先组类号，尽可能地加以分解，使之成为独立的检索点或成为能用于后组检索的限定因素。实现这个目标有以下几种途径：

(1)把自然科学以及部分社会科学的专论性类目细分，改为与相关的总论性组配标引。

这里包括四方面的内容：一是取消了专论性类目依专类复分表细分的方法；二是取消了专论性类目仿"一般性问题"细分的方法；三是使用"临近类目仿分"的类目，如果该类系还有与仿分对象相同类目结构的总论性类目(或"世界××"的类目)时，也改为主类号与相关的总论性类目组配标引；四是任何专论性类目的细分，都可以用与相关的总论性类目组配标引的方法处理。

凡自然科学各类"依专类复分表分"的，均改为与有关的总论性类目组配标引。各分类号分别著录在690字段。例如：

总论性类目

TH2/6 可依下表分。	TH 机械、仪表工业
01 理论	11 机械学
02 设计	12 机械设计
03 结构、零件	13 机械零件与传动装置
……………………	……………………………
06 制造工艺	16 机械制作工艺
07 运行与维修	17 运行与维修

TH2 起重运输机械

21 起重机

22 运输机

………………

依原规则，"起重机的结构"分类号为 TH210.3。

现改为：编制手工检索工具"TH2/6 仿 TH11/18 分(均加 0)"；编制计算机检索工具，主类号与 TH11/18 有关类目组配标引。起重手检目录"起重机的结构"，分类号仍为 TH210.3。机检数据标引为：

200 $a 起重机的结构 $f 黄益春著

690 $aTH210.3

690 $aTH11

905 $b0023415 $dTH210.3 $eJ504

这里，检索用分类号与排架用分类号相分离，以符合手检标引要求的分类号为主要分类号，著录在第一个 690 字段(主要是为有关用户选择)，排架分类号的确定一般以符合手检要求的分类号为依据，著录在 905 的 $d 子字段。

(2)把主类号使用通用复分表子目复分，改为主类号与通用复分表中的各主题要素组配标引。

由于《中图法》各通用复分表的子目均有前置标识符，因此在机读目录中可作为独立的检索点使用。凡主类号需用通用复分表复分的，均改为组配标引，即将通用复分表的子目号连同前置标识符，著录于 690 字段的 $ x 子字段。这里启用 $ x 子字段，用于著录形式复分的子目号，主要是为了某些手检与机检并用的单位在生成卡片式目录时，从排检项中过滤掉各通用复分表的子目，如果不编制手工检索工具，则可单独著录于 960 字段的 $ a 子字段。

在 690 字段标引时将用特定组配符号连接的复分表子目作为一个独立的标引单元和检索标识来使用，这在手工标引和编制手工检索工具中是不允许也是无意义的，但在计算机编目和检索的条件下，不但可以实现，而且赋予了新的功能。例如："英国十七世纪造船史"，手工标引为 U66-09561，使用这个分类号只能从造船工业入手进行检索，而"英国"、"十七世纪"、"历史"均不能作为检索入口。

机读数据标引为：

200 $a 英国十七世纪造船史 $f(英)史密斯著

690 $aU66 $x-09 $x(561) $x=332

905 $dU66

由于复分子目号前均有特定的组配符号做区分标识，因此在计算机检索中均可作为独立的检

索标识参与组配检索。例如用“U66 AND =332”组配，可以检索出“十七世纪造船工业”的文献。这样一来，原来凝固在先组类号中的主题因素便被离析出来，成为可独立使用的检索点。

(3)凡属多重列类的类目，均使用多主题要素组配标引。

多重列类法如果不与多主题要素组配标引结合起来，其优点基本得不到发挥。这是因为当文献主题只是从某一种属性进行论述时，可以用相应的类号标引，但当文献主题同时涉及几种属性时，只能依重点分或按“最前编号法”、“最后编号法”归入在前或在后的类目，这时提供的检索点只有一个。即使采用传统的主类号组配标引，其效果也是有限的，以“陶瓷卷式移相电容设计”为例，手检系统标引为 TM534.102:TM532.1:TM531.1，如果类号不轮排，检索途径不会增加，后面的类号只能起细分或说明的作用；如果采用类号轮排法，在手检工具中一是难以实现，二是也仅起到互见分类号的作用，不能用于组配检索。

现在规定进行多主题要素组配标引，即：

200　$a 陶瓷卷式移相电容设计 $ f 黄解春编著

690　$aTM534.1

690　$aTM532.1

690　$aTM531.1

690　$aTM502

905　$dTM534.1

这时用户就可以根据需要从任何角度进行后组配检索了，如用 TM502 and TM531.1 就可以检出“移相电容器设计”的文献，用 TM534.1 and TM532.4 就可以检出“陶瓷移动式电容器”的文献。

4.2　改造“能入专论的不入总论”的基本分类规则

“能入专论的不入总论”，是使用《中图法》分类的基本分类规则，强调在一定的学科、专业范围内按事物或问题逐次集中文献，这条规则在编制手工检索工具时是“金科玉律”，必须严格遵守，否则《中图法》按学科、专业集中文献的基本原则就遭到了破坏，其系统检索功能也不能实现。在机检系统中，由于不涉及目录卡片的多少、分类款目的轮排，所以，在有关主题标引专论性类号的同时，还应有选择地再标引相应的总论性类号。实际上，我们前面的几条规则已经部分解决了这个问题。除此之外，像某种理论、工艺、设备的总论性类目与其应用的关系等，除标引专论性类号外，还应该再标引相应的总论性类号。例如：

TM303.2	电机导电部件	总论性类目
TM382	交流微电机	总论性类目
U665.11	船用电机	专论性类目

“船用交流微电机导电换相器”，手检工具标引为 U665.11。机读目录应标引为：

200　$a 船用交流微电机导电换相器

690　$aU665.11

690　$aTM382

690　$aTM303.2

该条规则对机检系统十分必要，这是因为现在计算机检索用户已基本不依靠图书情报专业人员作中介，而是直接进行检索，他们对《中图法》总论、专论归类的复杂规则根本无从掌握(也无须掌握)，这样势必造成漏检；而且，在“交流微电机”类下查全各种交流微电机的文献(分散在各类)，也

是一种重要的检索需求。另外，改造后的标引方法为机检组配检索奠定了基础。

4.3 交替类目与使用类目同时标引

在编制手检目录时，使用类目与交替类目只能择一而用，在机读数据标引时则可同时标引交替类号(应去掉“[ ]”)，这样使具有多学科属性的主题分别在各相关学科、专业得到揭示，更能体现体系分类法按学科属性集中文献的要求，有利于用户按各自对某主题学科属性的理解检索文献。例如：

[X33] 环境卫生与卫生工程

环境卫生宜入 R14，卫生工程宜入 TU993。

“美国环境卫生现状”，机读目录可标引为：

200 $a 美国环境卫生现状

690 $aR14 $x(712) $x-1

690 $aX33

905 $aR14

4.4 对多主题、多属性的文献要充分揭示

在编制手工检索工具的分类标引中，由于受多种因素的制约，多主题、多属性的文献一般只按主要的主题、主要的属性进行标引，使文献中许多有价值的情报内容得不到揭示，因而也就无从检索、利用。在机读目录中我们就可以比较容易地对文献进行深标引，特别是对期刊论文、科技资料等尤为必要。虽然这样将增加文献标引的难度和工作量，但是我们不应忘记文献标引的目的、归宿就在于文献的检索。例如：

“异步电机无功就地补偿容量匹配”，手检工具一般只标引为 TM343 (异步电机)，机读目录还应标引 TM714.3(电力系统无功率补偿)、TM761.1(电力系统无功率自动调整)。

4.5 关于 690 字段的完善

CNMARC 中 690 字段只设有 $a(分类号)、$v(版本)两个子字段，这对于分类数据的标引和检索都是不方便的。由于分类号只使用一个子字段著录，在设计编目和检索程序时，由于不能对功能不同的子字段加以区分，很多功能将无法实现。为解决这个问题，应当根据编目和分类检索的需要，完善 690 字段的子字段设置。下面是一个完善 690 字段的设想。

$a 主类号

$d 用于组配的分类号、次要分类号

$e 交替类目类号

$f 总论复分号

$g 地区复分号

$i 国际时代复分号

$j 中国时代复分号

$k 种族民族复分号

$m 通用时空复分号

$v 版本区分

$w 版别区分

各子字段除了 $a 外，均可重复。其中：

$a　用于著录主类号，该子字段著录符合编制手工检索工具使用的分类号（即使用传统的标引规则进行标引）。该分类号标识文献的主要主题。该分类号在集中编目数据、商品数据中可供手检与机检并行的单位识别和选用。

$d　用于著录除主类号以外的其他分类号，如用于组配的分类号、次要主题的分类号、分析主题的分类号、并列主题的分类号等，实际上该子字段的各分类号即可以作为独立的检索点，又可以作为组配分类号使用。

$e　用于著录交替类号，增加检索途径。

$f/$m　用于著录各通用复分表的复分号，这部分复分号均可在组配检索中作为限定主题因素使用，从而提高检准率。当需要生成印刷型目录时，可通过子字段的控制，不出现在分类排检项。

$v　版次区分，用于区分某一版别的不同版次，如第三版、第四版等。

$w　版别区分，用于区分《中图法》的不同版别，如简本、儿童版、专业版等。

5　对类目结构和标引规则改造后的效果分析

5.1　基本兼容原有的分类号

由于将原来的仿分、复分改为与相关的类目组配标引，主类号选取的方法未变，原标引的分类号与新标引的机检分类号其中的一个仍保持前方一致。这种分类号的兼容性不影响原有数据资源的使用。排架分类号与检索分类号的分离，使得文献粗排架成为可能，如果排架分类号只用分类表列举的类号（不复分、仿分），将会提高对分类法修订的适应性，也会使检索分类号的编制具有更大的灵活性。

5.2　简化分类思路和复杂的组号技术

使用《中图法》进行分类标引，不但要把握类目的体系结构、辨识类目的涵义，还有掌握一套复分和组号技术（特别是加 0 技术）。经过改造的分类规则放弃复杂的类目复分、仿分、组号技术，直接使用多个相关的类号进行后组式标引。标引时，只须把握基本的“事物（学科、问题）——方面（性质、过程等）”或“事物——局部（部分、分支）”的要点即可。由于标引的均是一个一个独立的类号，只考虑选取的类号是否恰当，组配是否合理，而复杂的组号顺序问题不复存在，这样既简明又实用，不论标引人员还是检索人员都能很好地掌握，从而使标引思路与检索思路很容易统一起来。

5.3　大大提高《中图法》多途径检索的功能

多途径检索功能一直是体系分类法的薄弱之处，在手检的条件下无法得到有效地提高。采用新的标引方法，由于原来的仿分、复分要素都成为独立的检索点；由于文献主题从多角度加以描述，标引深度增加，为用户提供的检索入口至少增加一、二倍。这样无疑会大大提高《中图法》的多途径检索功能。

5.4　实现了计算机后组检索

计算机检索系统的最大优点之一是可以进行组配检索，输出满足特定条件的文献集合。但计算机组配检索的前提或基础是标引，只有对文献主题加以充分描述，检索系统才能分别抽取相关的

主题要素进行检索条件的匹配。由于我们把原来"焊接"到先组类号中的各主题要素加以分离,并增加了其他属性的描述,使计算机组配检索成为可能。不仅如此,这种主题要素的分离,也为分类号与主题词、著者的组配检索,创造了更大的空间。例如:

TM38 and V4	可检出航天类中有关微电机的文献
TM307 and (TD+P6)	可检出矿山和地质用电机维修的文献
K86 and 郭沫若	可检出郭沫若关于中国考古方面的所有文献
K and 契丹	可检出有关契丹人历史、考古、风俗的文献

主题法从事物名称入手检索,可检全该类事物的文献,但要限制在某一专业领域就相当困难,如果增加辅助的限定词检全率就会降低。体系分类法便于按学科或专业检全相关文献,但在某一学科或专业范围内检索特定事物的文献则力不从心。如果根据不同的需要使用不同的检索点进行联合组配检索,可充分利用各种检索途径的特长,弥补各自的弱点,达到若从主题词"电动机"入手检索(无"船用电机"一词),会检出大量非船用电机的文献;如果用"电动机"和"船舶"组配检索,因为很多文献是用"船舶系统"、"船舶辅机"、"航海仪器"等主题词与"电动机"组配的,因此会产生大量的漏检。

若用 U665.11 检索,只能检全"总论船舶用电机"的文献,而各种船舶设备专用的电动机还存在于以下各类之中(是隐含的):

U661.74	船舶模拟设备用的电动机
U664.152	船舶核动力控制仪表用的电动机
U664.811	船舶启动系统用的电动机
U666.123	导航陀螺仪用的电动机
U674.76	潜艇专用电动机

……………………………………

但我们把分类法的系统检索功能与主题法的特性检索功能结合在一起,用分类号控制检索范畴(造船工业)、用主题词控制检索对象(电动机),不论从检全率还是检准率来说都会得到提高(上述方法必须辅以上位词标引)。特别是当分类法中某些事物的属性没有作为划分标准使用(不可能完全做到这一点),也就是没有列出相关的类目时,完全可以通过分类号加主题词限定组配检索提高检准率。

5.5 解决或缓解了"集中与分散"的矛盾

《中图法》类目的基本结构是各类逐级划分为总论与专论两大块,文献标引的基本规则之一是"能入专论的不入总论",这样必然在总论性类目下检索不到专论性文献,在专论性类目下检索不到总论性文献。由于一般的用户极难掌握《中图法》复杂的总论一专论划分及标引规则,产生漏检是不可避免的。另外,在总论性类目下检索到相关的专论性内容,如在"汽车发动机冷却系统"类下,检索到各种汽车发动机冷却系统的文献,是一种重要的情报需求。经过上述改造,这个问题在自然科学类已经得到解决。例如:

U44　桥涵工程
441　结构原理
442　勘探设计
443　桥梁结构
.1　桥基
.2　墩台结构
.3　上部结构
.35　主梁
445　桥梁施工
.4　施工技术
.43　冬季施工
……………………
——总论性类目（原不含具体桥梁的主题）

以下各类可仿U44分。如遇多主题因素的桥梁文献，入前面编列的类……

448　各种桥梁
.1　各种用途桥梁
.2　各种结构桥梁
.21　梁式桥
.211　桁架桥
.212　板梁桥
.3　各种材料桥
.33　混凝土桥
.34　钢筋混凝土桥
.35　预应力钢筋混凝土桥
.4　各种形式桥
……………………
——专论性类目（多重列类）

前面编列总论性类目，容纳有关桥梁各方面问题的文献；后面编列专论性类目，按事物逐次集中该事物各方面的文献。按照传统标引规则，能入专论的不入总论，能入具体事物的不入其“方面”问题。这样，“预应力桁架桥主梁冬季施工手册”一文手工标引应为：U448.211.543。由于使用“最前编号法”的规则，在 U448. 35 预应力桥类下检索不到该文；由于使用能入专论不入总论的标引规则，在“冬季施工”类下也肯定检索不到该文（尽管这个类号中包含“桥梁冬季施工”的主题因素，但因“543”只是复分号，不能作为独立的检索点）；由于复分层次的限制，“桥梁主梁”这个主题因素在类号中也体现不出来，自然在“主梁”类下也检索不到该文。给用户提供的唯一检索入口是“梁式桥—桁架桥”。按改造后的标引规则可标引为：

690 @aU448.211
690 @aU448.35
690 @aU443.35
690 @aU445.43
690 @a-62
905 @dU448.211@eB346

这样，用户不论从"桁架桥"、"预应力桥"，还是从"桥梁主梁"、"桥梁冬季施工"类下都能检索到该文；而且，使用 U448.211 AND U443.35 组配可以检索到"桁架桥主梁"的文献、使用 U448.35 AND U445.43 组配可以检索到"预应力桥冬季施工"的文献、使用 U445.43 AND －62 组配可检索到"桥梁冬季施工手册"的文献，等等，使得原本在分类法体系中被分散了的事物，在机读数据中可以按任何属性得以集中，比如在"冬季施工"类下，不但集中了总论冬季施工的文献，也集中了专论各种桥梁冬季施工、各种桥梁部位冬季施工、各种桥梁施工设备冬季施工的文献，不论从"事物"还是从"方面"入手、不论从"总论"还是"专论"入手，都能检全相关文献。

5.6 提高检索效率

在手工检索工具中，从检准率来看，尽管类号可能相当专指，但组号的方法相当复杂，不是一般检索用户所能准确掌握的，而绝大多数只依靠已列举的类号进行检索，因此长而复杂的类号对检准率的提高，很大程度上是检索语言编制者理想的愿望。从检全率来看，按学科、专业集中文献的体系分类法，在按学科检索文献时应有较高的检全率，但由于事物的多学科属性在分类法中是强制归入某一学科的，在标引时又难以揭示主题的多学科属性，加上"集中与分散"的矛盾，就使检全率大打折扣。

改造后的类目结构和标引方法，能从多方面揭示主题的属性，提供更多的检索入口，实现后组检索，基本解决了集中与分散的矛盾，因此检索效率将会明显提高。

6 问题讨论

《中图法》改造中最大障碍是：如何处理社会科学门类中在较高层次按国家区分，继而再按其他标准细分的类目。如果不将这些类目展开，其体系结构得不到体现；如果按老办法层层复分，又解决不了"集中与分散"的矛盾，不能实现多途径检索。可采取根据具体情况进行分析处理，一部分保留原有的类目结构和复分的方法，如各国哲学、各国历史（重要的国家）；一部分可用地区复分表复分至国家或地区，再行细分则用相关的类目组配标引，如各国文学；多数类在特定的类目下，可使用地区复分表的子目及相关的类目组配标引。这样，各通用复分表的类目除上述情况外都用于组配标引，而不是传统的复分。

按上述思路对体系分类法进行改造，还存在如下问题，值得进一步研究。

（1）在社会科学有关类，如果在国家或地区区分之后（中国除外）不再仿照复分，进一步的区分被一个具有该属性的主类号组配所代替，将会削弱该类的系统性。

（2）由于主类号组配技术的广泛应用，一部分类号在末端失去等级性，对扩、缩检功能可能有一定的影响。

（3）由于文献排架实行粗分类，在开架借阅的条件下，不能形成对读者最有利的排列顺序（相对细分类而言）。

参考文献：

[1]张琪玉．情报语言学基础．武汉大学出版社，1987
[2]中国图书馆分类法（第四版）．北京图书馆出版社，1999
[3]中国图书馆分类法（第四版）使用手册．北京图书馆出版社，1999
[4]陈树年．在计算机编目和检索条件下对"组配编号"和"多重列类"的再认识．中国图书馆学报，1998(5)

[5]陈树年.文献标引和文献检索的新思维.大学图书馆学报,1999(1)
[6]陈树年.消除或缓解体系分类法中“集中与分散”矛盾的探索.图书情报工作,1998(11)
[7]陈树年.基于计算机检索对体系分类法改造的探讨.图书馆建设,1998(9)

文献分类法在网络资源组织中的应用

北京大学信息管理系　马张华

[摘要]本文对文献分类法在网络信息资源的组织的应用进行研究，对网络分类检索系统的类型、特点、建设方法，以及网络环境下自动标引技术进行了探讨。

[关键词]网络信息组织　文献分类法

有效组织和揭示网络信息资源，是合理开发利用网络资源的必要手段。如何将文献工作中长期发展起来的文献分类系统用于网络信息资源的组织，是各国文献工作者一直关注的一个问题。近年来，各国文献工作者经过努力，逐步编制出了一批以文献分类法为工具的网络资源检索系统。随着人们对网络资源开发利用的重视，目前这类系统的数量正在增加。因此，分析现有依据文献分类法建立的网络检索系统的类型、特点，并对有关问题进行探讨，是有益的。

本文研究的对象，为依据文献分类法建立的网络资源检索系统。主要根据"组织网络的类表"[1]和"电子分类表"[2]两个站点提供的线索，并结合对有关中文站点的了解进行。限于资料掌握的局限，不当之处，在所难免，敬请同行指正。

1　网络分类检索系统的类型

按照检索系统涉及的领域以及文献分类法使用特点，目前依据文献分类法建立的网络检索系统大致包括以下几种类型：

1. 依据综合性类表建立的综合性网络检索系统。这是现有分类网络检索系统中出现最多的类型。这类检索系统一般以现有的综合性文献分类法为依据，包括美国的《杜威十进分类法》(DDC)、《美国国会图书馆图书分类法》(LCC)、欧洲的《国际十进分类法》(UDC)、我国的《中图法》、荷兰的《荷兰概略分类法》(DBC)等。其中使用最多的是DDC，依据它为工具建立的综合性网络检索系统已超过10个。下表为网上部分比较典型的综合性网络分类检索系统：

	分类网络系统名称	编制机构	分类工具	其他
1	Net First	LCLC	DDC	约55,000
2	BUBL LINK	英联合信息系统电子信息委员会	DDC	约10,000
3	加拿大主题信息系统 Canadian Information by Subject	加拿大国家图书馆	DDC	收入有关加拿大的网络资源
4	NISS information Gateway	英国联合信息系统	UDC	
5	网络数库(Cyber Stacks)	美衣阿华州立大学	LCC	
6	荷兰电子主题服务系统 (DUTCH ESS)	荷兰国家图书馆等	DBC	协作编制
7	中国教育与科研网	中国教育系统	中图法	

2. 依据专业类表建立的专业性网络检索系统。比较典型的如，瑞典技术大学图书馆建立的“瑞典工程电子图书馆”(EELS)，该系统依据《工程信息分类法》(EL)，收入工程技术领域的1300多个经过选择的网上资源；又如，美国数学协会的“数学主题分类资源”，该系统依据《数学主题分类法》(MSC)，组织与数学领域有关的网络站点和网址，包括电子期刊，预印出版社等。

3. 依据综合性类表的某一专类建立的专业性检索系统。这类系统一般选择综合性类表的部分门类为依据，组织专业网络资源。如，加拿大多伦多参考图书馆的“业余爱好者天文学分类检索系统”，基本上是一个依据DDC的天文学类目建立起来的，有关天文学的网络检索系统。又如英国高教系统图书馆的“艺术、设计、建筑和媒体检索系统”(ADAM)，收入艺术设计、建筑和媒体、博物馆等领域的有关站点，依据DDC的700(艺术类)加以组织。

4. 建立与分类表相联系的主题范畴系统。这类系统以主题范畴分类为主，同时又通过标注分类号，保持与有关分类体系的联系。如美国加利福尼亚Pacific Bell建立的”兰网主题范畴”(Blue Web'n Content Categories)，以主题范畴为中心建立等级体系，主题范畴后附有相应的DDC号码；各层次的范畴类目均于相应层次中按字顺排列，其主题范畴的范围与DDC原类目范围往往并不完全一致。如下例即为按原有英文次序排列的该系统范畴表的局部：[3]

艺术(#700)
商业(#650)
英语(#800)
外语(#400)
健康与体育(#610)
历史与社会研究(#900)
　　人类学(#301)
　　加利福尼亚历史(#979)
　　时事(#070)
　　经济学(#330)
　　地理与文化(#900)
　　政府(#320)
　　人权(#361)
……………

5. 同时使用分类表和主题代码系统加以组织。如美计算机研究文献中心建立的“计算机研究资源”(CoRR)，同时采用计算机机械协会的《计算机分类表》和一个概略的字顺主题系统组织网络资源，两者均可以作为网络资源的浏览依据。又如我国教育站点的信息检索系统同时提供普通分类、学科分类、图书分类三个平行的分类系统，分别在各个分类体系下根据各自特点设置二级类目，每个分类系统都是一个独立展开的完整系统，用户可以根据需要选择使用。

在“网络组织的类表”目前列举的46个分类网络系统中，美国编制的最多，占17个；欧洲国家编制的总数为23个，其中英国占8个，德国5个。这一数字是与网络使用的普及程度相一致的。比较而言，我国的网络分类系统尚在起步阶段，随着网络使用的普及，文献分类法在网络资源组织中的应用必将逐步增加。

2 网络分类检索系统的特点

目前建立的网络资源分类检索系统不仅类型多样，在处理方式上也各具特色，下面从几个方面加以分析：

1.收入的网络资源大多是经过精选的。虽然现有的分类网络系统包括多种类型，但各类检索系统一般均有其明确的资源选择范围和质量标准。专业检索系统，如ADAM，EELS等只收入与其专业有关的文献、机构等的网址；以各种特定地区或对象为目标的专门检索系统，则收入有关该专门对象的站点，如"加拿大主题系统"只收录与加拿大有关的网络资源；以中学生为对象的"没有围墙的学校"只选择适合儿童使用的网络文献；即使是综合性的检索系统，通常亦根据其机构特点和用户对象的需要，确定一定的选择范围，按照要求进行选择。为了保证选择质量，不少检索系统均明确规定了系统的选择要求和质量标准，有的检索系统，如荷兰的DUTCH ESS还编写了专用的网络资源选择手册，[4]作为资源选择的共同遵守规范；与此同时，许多系统，一般还明确规定按一定的周期对该系统的资源进行查核，以确保链接质量，保证检索的有效性。

2.对类目进行必要的调整。由于文献分类体系并不是按照网络资源的情况编制的，在依据目前的类表对网络资源处理时，原有的类目并不完全适合实际使用需要，各种系统一般需要根据资源情况和系统的特点，采取相应的措施。大致有三种处理情况：其一，尽可能选用类表中接近或相关类目，只有在必要时才作少量变动。如："加拿大主题系统"在依据DDC组织资源时，基本上从原有体系中选择适用的类目，用于有关加拿大主题资源的标引；又如：Numeric List of BLZ - DEWEY - 1 Sections 和 Net Site by Number 等系统在进行资源组织时，一般均根据资源情况选用表中相应或相关类目，变动幅度较小。其二，重视根据网络资源情况对类目进行调整。如：BUBLE在按DDC体系建立检索工具时，强调以主题术语为基础进行标引，规定每个主题词下网络资源的数量大致应保持在5－15个之间，需要时可根据资源情况，于特定类下通过增加主题词加以标引[5]。如：该系统在"020图书馆学情报学"类下，就同时列出"图书馆学情报学综合资源"、"图书馆学情报学新闻和讨论"、"英国图书馆学情报学系"，"世界图书馆学情报学系"等多个类目[6]。其三，对类目设置及类名进行较大调整。如美Morton Grove公共图书馆的Webrary Links Menu在依据DDC的体系编制的同时重新命名了类名；又如"兰网主题范畴"等系统，基本上按资源情况重新设置类目。但比较而言，第三种情况所占比例较小。

3.对类目体系深度进行控制。为了使检索系统适合终端用户的使用习惯，多数分类网络系统十分重视检索工具的易用生，将类目体系展开的深度限制在一定的范围以内，使其较为简单明了。多数使用DDC的网络系统，将类目的划分深度保持在三级以左右；使用LCC的系统，一般在两位字母标记后展开网络站点；至于以儿童用户为对象的"没有围墙的学校"等，更是直接在基本大类之后列举网络站点；对于部分标引较专指的系统，则通过改进显示形式的方法予以解决。例：BUBL的分类深度有时可达到5、6级，该系统通过改进编排、减少显示屏数的方法，除总论、社会科学、自然科学、历史地理等门类的展示使用三屏外，其他门类只使用一屏即可以揭示所有类目。此外，一些系统通常不通过在屏幕上显示其上位类等级的方法揭示细目的位置，改进对类目体系的易用性。

4.加强类下说明。为了便于用户把握类目范围和资源特点，不少网络资源系统采用加强类下说明的方法。最常用的是于类下列举该类所包含的主题。这类列举不同于显示下位类，而是通过简练的术语进一步明确其范围或某些不易判断类目的归属。这一方法在网络工具中使用较为普遍，可以在一定程度上改进传统类表某些类目过分概括造成的不足。下图即为BUBL类下说明的

例子[7]。

001 知识,人文科学和研究 　　包括:学术,人文科学,研究方法,研究基金 002 图书 　　包括:图书艺术,图书史 003 系统 　　包括:复杂系统,控制论,模拟 004－006 计算机与互联网 　　包括:计算机科学、数据处理、网络、程序 　　参见:软件文档,数学和计算机主题术语 010 书目 　　包括引用参考指南 020 图书馆学和情报学 　　包括图书馆主页,图书馆操作,数字图书馆。 …………

以此同时,分类网络系统一般均重视对资源著录的改进,以便利终端用户对资源情况的了解。如:BUBL 对每个资源的著录,包括内容简介、作者、分类号、资源类型、国别等;"网络书库"在每一资源下的介绍包括:内容要点、主页设置、检索途径、制作机构、国家及检索建议等;Netfirst 的著录项目包括:数据库号、题名、类型、出版者、内容性质、人口、范畴、提要、LC 标题词、DDC 号等。通过这样的著录,用户可以方便了解资源的特点,根据需要作出选择。

5. 界面显示形式多样。检索系统的界面显示是直接与使用效果相联系的。目前分类系统界面的显示形式,大致有以下几种:其一,采用图形形式,如:"没有围墙的学校"等。其特点是,界面生动,可以吸引用户,使访问者耳目一新,但如果处理不好,用户不容易把握,一般适合于划分层次较浅的系统;其二,以列表方式显示。分为直接显示和分屏显示两种。直接显示通常直接或于大类后完整显示类表,有的还包括资源的列举,形式直观,但往往只能显示类目关系的局部,也不便于类下说明的增补,似不宜于大型系统使用,如"加拿大主题资源"等目前采用这一方法;分屏显示层次清楚,也便于说明文字的增补;不少表为了指明其类目的等级,通常在类前标明其所属的上位类,便于识别与转换。为了减少显示的屏数,不少系统在三级以下不严格按类目等级划分屏数,以加强其直接性。对于网络资源的显示,系统大致采用两类方式:一类将界面分为若干部分,逐级揭示相应的的网络资源,类似传统的分类目录,例:Net First 的原型 Mr. Dui's Topic Finder 采用这一方式;另一类一般则在一定级别,如三、四级以下对资源加以显示,例:BUBL 等。

6. 多种检索途径结合。在采用分类体系进行系统显示的同时,上述系统一般还提供著者名、题名、类名或关键词、标题词等多种形式的查检途径。为了方便多途径检索,有的检索系统还明确规定"在可能的情况下,尽量采用受控词汇",以减少不确定的词汇,保持标引的均衡性和相关性,降低检索误差。例:BUBL LINK、加拿大主题系统等均规定其主题术语尽量依据 LCSH,后者同时还依据加拿大标题表等。与此同时,一些系统还采用提供多种浏览途径的方式,同时列举主题范畴代码,提供字顺浏览途径。如,美国计算机研究资源(CoRR)除使用 ACM 计算机分类系统组织外,还按照资源的主题领域加以显示;又如,我国教育网的检索系统采用了同时使用多个平行分类系统的方法。此外,不少系统还结合网络技术特点,在分类系统中揭示与其他系统有关门类或专门站点的联系,使该系统与已有的网络资源检索系统结合成一体,以便利用户对整个网络资源的查找。下表列出英国信息服务系统的 NISS Information Gateway 各类与相关专业系统或站点的联系:[8]

NISS 系统	相关站点	相关站点
0 总论	使用 ROADS 软件的 主题检索入口(全部)	BUBL - LINK(全部) PINAKES(全部)
1 哲学	网络上的哲学	
2 宗教、神学		
3 社会科学、法律、政府	SOSIG(社会科学)	Biz/ed(商业信息)
5 数学和自然科学	SciCentral	ChenDex(化学) PAW(物理)
6 应用科学、医学、技术	EEVL(工程)	OWNI(医学)
7 艺术、娱乐、运动	ADAM(艺术) RUDI(城市设计)	医学和人类学数据服务
8 语言、语言学、文学		
9 地理、传记、历史	IHR(历史)	

7. 处理对象层次多样。这里的处理对象,主要指系统收入的资源类型。包括以下两类情况、其一,同时收入多种资源类型,包括各类网络文献、电子期刊、连续出版物、会议、机构、软件甚至大型检索系统等。多数网络分类系统采用这一标引范围。其特点是涉及面广,可以通过它进入各种类型资源;不足是,一些机构或检索系统往往有多种功能,用户进入后需要重新开始一轮新的查找,容易迷失方向。虽然一些系统在著录中加强了对资源类型的揭示,但随着资源数量的不断增加,这一形式很难适应使用需要。未能以专类、特别是复分的方式充分揭示资源类型,是这类系统的普遍不足。其二,只收入特定的资源类型。这类系统一般根据使用对象的需要和系统的性质,明确规定标引的类型范围。如,美密希根州的"网络联机文本图书馆",只收入文本文献;美威尔斯大学的"PICK 系统",只以文本文献、报告、会议录、档案等为标引对象;两系统均不收入机构、活动等其他类型站点。又如,以儿童为使用对象的"没有围墙的学校",多数类下只收入经过选择的一种资源,如,其传记类只介绍一种传记工具书作为阅读对象。这类系统用户查检时针对性更强,但由于其链接范围受到限制,通常适合于在一定范围内使用。

8. 编制方式多样。现有的网络分类系统多数为公共图书馆或专业图书馆编制,除个别系统依靠网络爱好者的业余活动建立外,通常由专业人员负责资源的选择和处理,从而使质量得到一定的保证。多数系统依托某一机构进行,并有明确的分工;与此同时,也出现了由多个机构协作编制的检索系统。比较典型的如荷兰的 ESS,就是荷兰国家图书馆与众多的大学、公共图书馆在馆际合作的基础完成的。各馆均负责一定的主题领域,并遵循共同的编制规则,从而使系统保持在一定的质量水平。有的网络系统虽然是由单一机构完成的,但在编制过程中注意了与该系统其他部门资源的协调,在一定程度上也具有协作开发的性质。如:BUBL LINK,ADAM 等均属于此类情况。许多系统在分类系统建立的同时,通过链接的方式建立起与相关系统的联系,促进了系统之间的结合。为获取用户需要的高质量网络资源,各种检索系统一般均十分重视用户的推荐,使系统能具有较强的开放性,能够根据用户需要不断加以改进。

3 对有关网络分类检索系统若干问题的讨论

众多分类网络资源系统的出现,是各国分类界努力探索的结果,同时也表明人们对分类系统作用的承认。但也应该看到,将文献分类系统应用于网络资源组织,同时也是对传统分类体系的一个挑战。与原有文献分类检索系统相比,处理对象上,两络系统涉及到各种网络资源,数量大、种类

多、动态性强、对类目系统的要求更高；使用要求上，网络系统面对的是广大终端用户，检索需求千差万别，并要求系统有更强的易用性；技术环境上，网络系统是与各种计算机网络的最新技术直接联系的，要求分类体系能适应技术发展的需要。如何根据网络系统的特点和要求，对分类体系加以完善，改进标引方法等，是建设网络分类检索系统中必须面对的问题。本文谨对下述问题进行概要讨论：

3.1 现有分类体系的完善

目前使用的网络分类系统，大多对原有类目体系进行了一定程度的调整，这是网络检索系统的特点决定的。传统分类工具是根据书目文献对象和用户使用需要决定的，网络系统既然在处理对象和用户需求方面均存在着巨大差异，自然要求对类目体系作出相应变动。因此，如何根据网络检索系统的特点调整类表，使分类体系更加适合使用要求，是网络检索系统建立过程中必须解决的问题。关键在于如何了解调整的规律，使其尽可能合理。目前对类表的变动包括：对类目进行增补，改变类名，控制类目等级，局部调整，增加类下说明，增加入口词等形式。但各个系统在变动形式、调整幅度等方面，方法多样，特点不一，有的任意性比较大。要避免任意性和盲目性，有效进行分类体系的调整，就必须了解调整规律，至少应对下述问题加以探讨：

①研究调整类型。包括探讨调整可能采用的形式，分析各种调整形式对检索效果的作用和与网络环境的结合效果，了解他们对原有系统的影响等，以便弄清各种调整形式的使用特性。

②确定调整依据。目的是增强调整的规范性，使网络类表的调整有规律可循。应尽可能确定具有可操作性的依据。如，可选择以网络资源数量和用户检索词等作为调整的依据；同时对影响使用效果的各种因素进行分析，发现调整规律。

③完善调整机制。目的是通过对调整过程的规范对调整质量进行控制，如，指定专人负责类表的调整，明确调整的程序，加强对调整的研究和管理等。

对传统文献分类法的调整和管理，文献分类界进行过许多研究，积累了许多经验，如果能把这些研究成果与网络资源处理的特点相结合，就可以有效促进网络分类系统的完善。

3.2 现代分类技术的使用

目前网络分类体系中使用的基本上都是传统分类技术，特点是直观、简明，易于使用，用户可以利用分类体系的展开进行浏览；其不足是，未能将网络检索的特点与本世纪来分类研究的成果应用于网络系统，缺乏根据网络资源的特点进行揭示的能力。虽然这些系统一般都同时提供多种检索入口，如国别、标题词或自由词等，但在多数系统中，各种形式之间是各自独立的，不能结合进行；此外，主题词的数量不够充分，主题途径的显示也比较简单，从而在一定程度上限制了分类体系能力的充分开发。因此，如何根据网络资源的特点，合理引入分类研究的成果，十分关键。我认为应逐步在分类网络系统中引入和加强下述技术：

▲组配技术。根据网络资源的特点，选择适合的分面，供组配使用。如：可以根据使用需要，为网络资源类型、国别等建立供组配使用的窗口，同时将某些途径如主题词等的独立检索变为可以结合检索，从而增强资源选择能力。

▲分类主题之间的转换。目前的许多系统，分类浏览与主题浏览之间的联系不够密切，应增强两者之间自由切换能力，便利从主题的角度充分利用分类系统。

▲主题结构显示。主题法以事物对象为中心集中相关材料，是对学科分类的一种补充，目前的主题浏览过于简单，未能对主题对象之间的联系进行必要的揭示；应在主题浏览中引入对主题结构

的显示，满足从事物对象角度进行浏览查找的能力。

▲增加自然语言语词入口。许多系统重视主题词的规范，但对增加自然语言语词，建立这些语词与规范词之间的联系重视不够，可在用户检索词的基础上进行选择，不断充实自然语言中的词汇，使用户可以从不同的语词出发利用系统进行检索。

当然，在现代分类技术引入时，应注意其使用的合理程度，使其能在保持现有系统的简明性的同时，增加分类体系组织和揭示文献的能力。

3.3 探索网络系统建立的方式

如同传统的文献检索系统一样，由于不同层次、领域、地区的网络用户对检索存在着不同的要求，编制特点各异的分类网络系统是必要的；但是网络空间与传统文献检索系统在环境和技术上存在着不同，如何根据网络环境的技术特点，在网络系统的建立过程中采用相应的编制方式，是一个有待进一步探讨的问题。

与传统的检索系统相比，网络检索系统面对的环境更加开放，可以采用的方式更为灵活，不仅其处理的资源对象是共有的，而且通过链接的方式可以方便地进行相关站点之间的连接，各个检索系统之间的协调十分便利。因此，应根据网络环境的特点探讨其适用编制的方法。从现有系统建立的情况看，网络检索系统至少存在着以下几种编制方式：

①选择。即根据网络特点，从现有网络系统中精选需要的网址，建立供特定对象使用的小型系统；

②补充。即根据对资源情况的掌握或了解，只编制其他分类系统没有收录的资料，作为已有系统的补充；

③借用。即通过链接，将有关系统作为该系统使用的相关站点，指引用户直接使用；

④协调。即一系统的不同部门之间根据各自的特点，分别编制相应的目录，可以在一定程度上避免不同部门之间的重复劳动。

⑤协作。即数个机构根据需要分工合作，依据统一的规范建立同一系统。

还可能有其他可供采用的方法。上述方法中，虽然某些方法在传统系统的建立中也有使用，但其特点是不同的。因此有必要系统分析网络环境对分类检索系统编制的影响，了解网络环境中各种可以更加合理地组织资源的编制形式，探讨各种编制方式的技术、规范以及相关问题，包括其中的知识产权问题。检索系统的建立也是一种智力劳动，应当探讨各种编制方式在网络环境下使用的合理界限以及应注意的问题。

3.4 关于标引自动化

对于标引自动化，本文只简单强调以下几点。

①自动标引是网络资源标引的发展方向。采用分类的方法组织网络资源，用户可以按照系统的展开进行浏览，根据内容之间的关系扩大或缩小检索范围，方便、直观，并具有一定的启发性。但目前的网络标引是由人工进行的，不仅成本大，费时长，而且收录的资源范围也会因标引速度受到影响，很难适应网上信息组织的需要，与网络环境的要求是不匹配的。可能的解决办法，是采用自动标引或人机结合标引的方式，即由计算机进行标引，由人工进行质量控制和调整。由于网络资源数量大、范围广、动态性强，因此，与传统文献处理的需要相比，网络检索系统对自动标引的需要远为迫切。

②网络标引自动化将是对自动标引已有研究的推进。根据汉语的特点，我国的标引自动化研

究包括:语词的切分和识别;主题概念的分析和提炼;自然语言向受控语言的转换等内容。其中,进行分类转换的关键之一,是应当建立起一个完备的语词一分类体系的转换系统,它可以采用类似分类索引的形式,也可以是主题词、分类法的对应表。但不管是何种形式,除需要根据网络资源的特点对类表进行完善、调整,大量充实入口词外,还应解决复合主题的转换问题,包括如何在现有基础上进行引用次序的调整、使类表具有较强的规律性,并能够根据不同主题概念的特征,按照一定的次序加以组配等。必须结合网络标引的特点,在现代分类理论和技术的基础上进行研究和探索。

③自动标引是能够成功的。虽然标引自动化目前仍处于摸索阶段,但应该看到,计算机的使用至今不到60年,从50年代关键词标引系统,到目前一些逐步投入实际使用的人机结合的系统,已经在自动标引的许多领域取得巨大进展。由于计算机技术的研究目前处于社会关注的中心,各国投入力量大,技术推进迅速,随着知识经济的发展,社会对信息处理的重视,文献自动标引必将取得较大突破。因此我国应重视对网络信息处理的研究,加强汉语自动化标引的探索。

参考文献:

[1]http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm

[2]http://orc.rsch.oclL.org:6109/classification

[3]http://www.kn.pacbell.com/wired/bluewebn/

[4]http://www.konbib.nl/dutchess/index.html

[5]http://bubl.ac.uk/link/about.html

[6]http://link.bubl.ac.uk:80/lis

[7]http://link.bubl.ac.uk/000

[8]http://www.konbib.nl/dutchess/docs/links.html

(本文所参考的以上站点中的资料截止到1999年5月6日)

网络环境下分类法的变革与发展

北京大学信息管理系　王忠红　张　涵

[摘要]本文从分析网络信息二次组织的现状入手,探讨传统分类法在网络环境下的应用与发展问题,通过非基于分类法与基于分类法两类浏览工具的比较,论证了传统分类法在网络环境下应用的可行性,并以论证为依据,对分类法在网络环境下的应用和发展提出建设性的发展策略。

[关键词]网络环境　分类法　万维网浏览工具

1　引言

分类法是图书情报领域经典的情报检索语言之一,以1876年DDC的问世为标志,分类法的发展已走过了一个多世纪,在这一个多世纪的发展历程中,分类法形成了自成体系的一套信息、知识组织方法,这套方法能不能适用以及如何适用于人类正在全面进入的一个新的信息环境——网络环境,将直接关系到分类法在21世纪的前途和命运,也将关系到21世纪的情报检索语言建设。

网络环境下的信息分类组织方法表现为列表浏览型搜索引擎(hotlists),它以超文本方式将信息按学科、行业或用户需求从宽到窄组织起来,供用户逐层浏览和查询,有的文献称之为范畴目录型搜索引擎或主题指南。由于目前大部分搜索引擎都同时提供关键词检索和列表浏览,为了加以区别,本文突出它的浏览功能,称它为浏览工具,万维网检索工具是目前最为流行和最为常用的网络检索工具,本文以万维网浏览工具为主要研究对象。

万维网浏览工具按研究需要可分为非基于分类法和基于分类法两类,非基于分类法浏览工具是指没有参考现有的图书馆分类法、设计者自我设计的;基于分类法浏览工具是指以图书馆分类法为浏览结构基础的。非基于分类法浏览工具可被视为是网络环境下应运而生的分类组织方法,基于分类法浏览工具可被视为是传统分类法在网上应用的尝试,本文试图通过两类浏览工具的比较来探讨分类法在网络环境下的应用和发展问题。

2　两类浏览工具的比较

非基于分类法浏览工具通常由计算机人员或网络爱好者设计,它以收集时事新闻、商业信息、政府信息和娱乐、生活信息为主,报据信息的热门程度和用户的爱好需要,将信息分成几个大类,每个大类下又分成若干个子类,类目之间按照等级关系排列,同一等级的类目按字顺排列,上下类目之间及与记录之间以超文本方式链接。该类浏览工具以Yahoo为代表。

非基于分类法浏览工具的主要特点是运用了计算机环境下的操作方式,通过多角度、多视角揭示和组织信息,来实现部分或全部多途径检索。在列类上,大量采用了多重列类法和双表列类法,由于非基于分类法浏览工具的同一等级类目按字顺排列,上下类目关系的逻辑性不是很强,所以它

的每一等级类目的划分几乎都使用了多个分类标准，如 Yahoo 对教育类二级类目的划分，使用了教育等级、教育方面、教育类型、文献形式等多个划分依据，典型的划分事物类型的多重列类法在 Yahoo 中也被使用，如 Yahoo 对图书馆类型的划分，使用了性质、载体、学科、地区、人群五个划分依据，Yahoo 还把多重列类法应用于整个类目的划分，也就是双表列类法，如 Yahoo 的历史类和教育类除了列举按历史和教育主题划分的类目外，还分别按地区、学科、年代和地区、文化或人群、学科划分出另外三个类表(见表 1)；在类目横向关系的处理上，Yahoo 采用了类目重复设置的方法，也就是对于涉及多学科或多主题的类目在多个类目重复反映，如艺术疗法入艺术和健康两个类，计算机公司入计算机与因特网、商业与经济两个类，教育理论期刊入教育类的教育理论和期刊两个子目等；在信息标引上，Yahoo 采用了互见标引法和分析标引法，多学科或多主题信息采用互见标引法，根据其涉及的主题归入相应的类目，综合性信息采用分析标引法。所以 Yahoo 中的很多复合主题都可以从几个检索途径来检索，比如关于"美国数学高等教育教学改革"的信息，既可以从教育类入手，也可以从地区类和科学类下的数学子类入手，在教育类下从高等教育、教育改革二个子目和地区教育浏览、学科教育浏览二个类表下都有可能查询到所需要的信息。

表 1　Yahoo 教育类和历史类的双表列类

Home:	Home>Arts>Humanities>
Education ————教育类	History ————历史类
Browse by Region(120)—按地区划分	Browse by Region(4668)—按地区划分
By Culture or Group(415)—按文化或人群划分	By Subject(4473)————按学科划分
By Subject(11)————按学科划分	By Time Period(1927)——按年代划分

非基于分类法浏览工具也存在不少问题，它的基本大类设置不太科学，如 Yahoo 的大类设置了人文科学类、社会科学类，却没有自然科学类；知识体系也不严密，如 Yahoo 经济类的各部门经济缺少通讯经济和工业经济；类目划分没有规律，如 Yahoo 的教育类既有多重列类又有双表列类，但却没有任何两个类列采用了统一的分类标准和引用次序；类目等级递进性差，如 Yahoo 的计算机类从大类直接进入"个人计算机"子目，中间省略了"计算机类型"这个类目；类目从属关系处理错误，如 Yahoo 的科学类把隶属于生命科学的生物学、生态学、古生物学和生命科学总论并列，社会科学类把隶属于人文科学的历史、语言学和人文科学总论并列，把隶属于历史的谱系学和历史总论并列，同类问题在 Yahoo 中很普遍；类目横向关系处理没有规律可循，如 Yahoo 中涉及两个学科或两个主题的类目是采用重复设置的办法，但它的设置规则却无规律可循。这些问题的关键在于缺乏专业理论的指导。

基于分类法浏览工具通常由图书馆界人员设计，以英国斯特拉斯克莱得大学安德森图书馆开发的基于 DDC 的 Bubllink 为代表，目前国际上已有不少分类法被应用为浏览工具的结构基础，如 UDC、LCC、NLM 等，其中以 DDC 的应用为最广，已被应用到因特网上十多个站点。[1]

基于分类法浏览工具的主要特点是基本上保留了分类法原来的知识体系和类目体系。在知识体系方面，虽然增加了一些网络信息新主题，但数量非常有限，仍以收集学术信息为主、面向教育科研人员；在类目体系方面，以分类法简本为浏览结构基础，仍然保留了分类法的单线、一维体系，由于类目级别不长，基于分类法浏览工具的类目级别一般是二、三级，多重列类法极少被使用，如对图书馆类型的划分，大多数基于分类法浏览工具无一例外地采用了学科性质作为划分标准；信息标引则采用整体标引法，一个主题一个位置。但它同时也保留了分类法的科学性和逻辑性，它的基本大类设置科学、知识体系严密、类目体系井然有序、类目关系明了、著录详尽。

从以上两类浏览工具特点的分析中已经可以看出两类浏览工具的异同，如以图表形式来表示

的话，则能更加全面和清晰地看出两者的异同(见表2)。

表2 两类浏览工具的比较

比较项目＼比较对象		非基于分类法浏览工具	基于分类法浏览工具
知识体系		知识体系不严密，能反映网络信息新主题	知识体系严密，反映的网络信息新主题少
类目体系	分类标准的使用	大多数等级类目的划分使用多个分类标准	大多数等级类目的划分使用一个分类标准
	类目纵向关系	类目从属关系不严格，甚至存在不少错误	用标记符号和排版格式表示，类目从属关系清楚
	类目横向关系	用类目重复设置表示，类目横向关系不清楚	用交替类目、参见、注释表示，类目横向关系明了
	类目排列	不同等级类目以超文本方式表示，同级类目按字顺排列	不同等级类目以标记符号和排版格式表示，同级类目按自然序列或具有检索实用意义的人为序列排列
标引		互见或分析标引，著录简单	整体标引，著录详尽
检索途径		部分或全部多途径检索	单途径检索
功能		浏览，扩/缩检(将关键词检索限制于某一范畴)	浏览，族性检索，扩/缩检
用户定位		普通用户	教育科研人员
技术利用		超文本技术利用得好	超文本技术利用稍逊一筹

从两类浏览工具的比较中可以看出，分类法在网络环境下仍然是有生命力的，尤其是分类法的一些传统经验与方法，如等级列举、系统排列、交替排列、相互参见和互见分析等方法，比如，Yahoo对类目横向关系的处理是采用类目重复设置的方法，该方法不仅不能清楚地说明主题之间的关系，还不能适用于两个具有交叉关系而不能形成独立类目的主题，而且会带来类表的体积问题，而分类法的交替类目和相互参见方法，不仅能够清楚地说明类目之间的关系，而且在超文本环境下是可以直接跳跃过去的，不会带来类表的体积问题。应该说，等级列举在网络环境下的发扬，正是证明了传统方法的生命力。但我们同时也要看到，将产生于传统信息环境、满足传统需要的分类法直接应用于网络环境，或者仅做一些小小的改动，是无法全面体现网络环境特征和满足网络信息需求的。就目前而言，基于分类法浏览工具的影响力和用户访问数是远远低于非基于分类法浏览工具的，尤其是Yahoo的影响，不少网络搜索引擎和电子产品都借鉴它的分类特点。分类法要对网络信息进行组织揭示，还需要发展和变革。这种发展和变革包括二个内涵，一个是对在手工操作基础上建立的传统经验与方法的发扬光大，另一个是对分类法不适宜网络环境的地方进行改革，包括对网络环境提供的技术与手段的利用。

3 网络环境下分类法的变革与发展

3.1 知识更新

分类法在传统信息环境下主要应用于图书情报机构，图书情报机构的主要职能是教育和为科研服务，所以分类法处理的对象以正式出版物为主，包括少量的“灰色文献”，不包括私人信息，从内

容上来看，囊括了人类的全部知识门类。网络的功能与图书情报机构的职能有所不同，它既有教育与科研服务功能，又具有商业、生活、娱乐及一般信息交流功能，而且以后者为重，加上网上发布信息自由而方便，无需经过任何权威机构的审核和认定，所以，网络信息既包括正式出版物，又包括半非正式出版物，如机构名录、工作进展报告、学位论文、讲课讲义、产品介绍等灰色信息，还包括非正式出版信息，如新闻报道、股票行情、体育赛事以及电子邮件、电子会议、电子论坛、电子布告板中的动态性极强的信息，从内容上来看，除了囊括学术信息的知识门类外，还包括非学术信息的娱乐、生活、商业类信息以及私人信息。

面对信息内容的扩充，分类法有两种选择，一是保留原有的知识体系，选择符合知识体系的信息，二是对分类法进行知识更新。第一种选择也不失为一种良策，基于分类法浏览工具以其严谨的学术性与非基于分类法浏览工具形成鲜明的对比，形成了自己的特色，但这种选择是一种较为被动的策略，况且目前基于分类法浏览工具中存在一个较为严重的问题是，不少类目后没有揭示的信息记录或只有一、二条记录。[1]所以，笔者认为对分类法进行知识更新势在必行。

从分类法在网上的应用现状来看，分类法的整个知识体系在网络环境下还是适用的，存在的两个问题是，一是分类法的有些主题没有可揭示的网络信息，也就是说网络信息的知识门类不如传统信息齐全，比如农业类、数学类、医学类、工程类等，俄国物理学家 Lebedev 认为，在因特网上发现的科学信息，大约是从高质量科学数据库获得信息的 10%—20%，[2]这个问题随着网络的发展，越来越多的高质量、学术性出版物上网，会逐步得到解决。二是分类法中没有与网络信息相对应的主题，需要作一些补充，笔者认为补充应该是有选择性的，网络信息良莠不齐，有害信息和无意义的信息完全没必要反映，从网络信息的内容来看，分类法需要补充反映的信息大致有以下四类：

——由计算机技术、网络技术等信息技术发展而带来的新主题，包括信息技术本身飞跃发展而带来的新主题和信息技术对其它领域渗透而带来的新主题，如 2000 年问题、虚拟现实、赛伯文化、虚拟艺术、电子出版、网吧、网上学校、网上商业以及网络指南等等；

——灰色信息，如公司概况、产品介绍、教育机构的课程设置与招生考试信息、教师讲课讲义、工程进展报告、内部出版物等等；

——动态性消息类信息，如新闻报道、股票证券行情、体育赛事等等；

——娱乐、生活信息，如旅游与休闲、购物与销售、游戏、饮食与住宿、医疗与急救、交通、天气预报、生活指南等等；

对于前面两类信息可在分类法中有关类目下增设，娱乐、生活信息可专门设一个类加以反映，与学术信息分开，至于变化较快、用户又比较关注的动态性消息类信息，可在类表体系以外，采用类似报纸的版面、版块、专栏的形式来加以组织与提供。

网络信息内容的改变不仅影响着分类法的知识体系，还影响着分类标准的选择。如对图书馆类型的划分，分类法一般以图书馆性质为主要分类标准，面网络环境下的图书馆概念被虚拟化了，不再是个实体，任何一个提供特色信息的网站都可自称为图书馆，对于用户来说，更关注的是图书馆的馆藏特色，图书馆的性质变得无关紧要，原来的分类标准就需要改为以学科为主要分类标准。

3.2 体系多维化

传统分类法的体系是一种典型的树型结构体系，它根据类目之间的内在联系，将类目从总到分、从一般到具体、从低级到高级、从简单到复杂逐级展开，这种体系非常符合事物发展的一般规律，也符合人们认识事物的思维习惯。但它有一个弊病，即是“以一维方向排列多维思想的物质形态”[3]，所有的主题只有一个位置，不仅要求用户与分类法逐级展开的思路达成一致，而且其间每

一步都不能出错，否则检索就不能成功。显然，这是与多维性的知识空间、多维性的检索思路相排斥的。

分类法的这种一维、单线的体系是手工检索环境下为满足排架需要而产生的，网络环境下电子信息的组织摆脱了图书排架问题，超文本技术为多角度、多侧面组织信息与多角度、多途径浏览与检索提供了技术支持，使分类法实现体系多维化有了可能。

分类法的体系多维化是指在保持分类法原有的系统性特性的前提下，实现多途径检索与多维检索。在类目的纵向关系上，仍然保持类目的层层划分与等级关系，浏览工具类目等级至少要达到五级，并且要强调类目的层次分明，以符合人们的一般思维方式，在类目的横向关系上，多建立类目之间的横向关系，传统的交替类目、参见与注释方法，在万维网环境下，既不影响类目的系统性，又能迅速指引用户到相关的主题，并且还不存在类目体积问题，应大量使用。同时还要多角度揭示信息，即使用多个检索点。具体措施如下：

——多角度揭示信息。即使用多个分类标准，包括多重列类法与双表列类法。为了不破坏分类法的系统性，多重列类法应限制在较低层次的类目使用，一般以在三级以下的类目使用为宜。双表列类法在超文本环境下可以大量使用，因为不存在体积问题，像 Yahoo 的双表列类，短短一个标题，既明了，又提供了另一个检索途径。

——大量建立类目之间的横向关系。对于与两个学科有关的主题，采用交替类目的方法；对于两个概念内涵交叉的主题，采用参见方法；对于一类事物的总论或专论，可以在总论下重复反映专论内容，也可以在总论下设置专论的交替类目。这些方法都可以起到指引用户到有关类目的作用，而且在超文本环境下，是可以直接跳跃过去的。

分类法的多途径检索的实现除了通过类目体系改造外，还需要标引工作的配合。对于多属性、多主题信息，传统的分析标引法、互见标引法都可以广泛使用，电子目录不会带来体积庞大问题。至于多重列类法的标引，可以使用多个类目重复标引，也就是类号轮排，以达到既能保证全面揭示信息，又能实现多途径检索的目的。

3.3 面向用户

分类法在传统信息环境下与在网络信息环境下服务的用户有所不同。在传统信息环境下，分类法主要应用于图书情报机构，图书情报机构有比较明确的用户对象，如大学图书馆以学生和教师为主。而在网络环境下，网络打破了信息获取与传递的地理障碍，与网络相接的任何一个终端用户都可以成为分类法的检索用户，用户成份变得多样化，尤其是易操作、界面友好的万维网的兴起，使得社会各行业、各阶层的人士都可能成为网络的使用者，不再仅限于计算机专业人员或科研、教育人员，普通用户占的比例越来越大。其次，在传统信息环境下，用户对分类法的使用往往是在专业检索人员这一中介用户的帮助下，或者直接由中介用户来完成，而在网络环境下，终端用户成为分类法的直接使用者。

除了服务的用户群体发生变化外，网络环境下人们的思想观念也发生了变化。在网络环境下，由于信息技术已发展到一定的水平，人们开始追求以人为本的技术设计思想，比较强调用户的利益，万维网信息服务本身就是一种以用户为中心的服务模式，万维网浏览工具的出现则是出于对初上网者、初学者的考虑。在网络环境下，对用户是否"友好"成为系统能否成功的关键之一，无论是检索系统还是检索语言，可近性、易用性都被排在效率与质量之前，一个难以接近、难以使用的系统，被认为无效率可言，无质量可言，[4] DDC 在网上被设计者青睐的这一原因就是它易于使用，BUBL 与 OMIN 两个网站原本是采用 UDC 为浏览结构的，后改用 DDC 与 NLM，主要就是因为

UDC 难以使用。[4]正如美国学者 T. Saracevic 所言："正在出现的新一代信息系统是走向成功还是失败，取决于能否极大地增进对用户依靠获取和利用过程的理解"。[5]

传统信息环境下的分类法在用户定位方面有二个误区，一个是强调类表的科学性和专业性，忽略类表的易用性，在类表的编制中往往请一些专家和学者参与，而实际上类表的使用者不可能个个是某一学科领域的专家，对于他们来说，短时间内熟悉高度科学化的知识体系难度太大；另一个误区是把中介用户——分类标引人员作为最终用户，忽略检索者的需求与利益，类表的设计者会吸取、接受分类人员的意见与反馈，为分类人员进行分类法新版使用培训，包括类表的电子化都是为编目员设计的，如 DDC 的第三代产品——视窗 DDC 适合编目员使用，却不适合在网上用，真正的最终用户——检索者却被放到了无足轻重的地位。总的来说，设计者强调标引的质量与检索效率，忽略了易用性，这是影响分类法使用的重要原因。

网络环境下的分类法要体现用户的利益，首先就要在观念上改变这二个误区，从面向中介用户转变到面向最终用户；其次是要在实践中贯彻面向用户原则，即实现界面的"友好"，具体措施如下：

——研究普通用户的一般思维方式、检索习惯和需求特点，按照普通人的思维方式、使用习惯与方法来设计分类法，包括知识体系的设计、分类标准的选择、类目概念的表达等类表编制的各方面，尽量减少用户的先期学习负担。这里所说的普通用户是指大多数用户，与分类法的科学性并不矛盾，分类法的科学性应该是符合大多数人的认知，而不是少数学术专家或权威。

——利用各种先进的检索技术，如超文本、模糊逻辑、人工智能等，为用户提供多样化的检索模式，使用户能用自己的术语和思维过程自由地完成检索，而不必被迫接受设计者设定的唯一检索模式，同时又能保证分类法作为受控语言原有的检索效率。

——将专业检索人员的中介工作由"前台"转向"幕后"，将专家们的研究成果和丰富的检索技巧转化为系统的功能与帮助工具，为用户提供必要的帮助和指导。

3.4 利用超文本技术

超文本是一种与传统文本相对的非线性信息组织方式。传统文本的信息组织方式是线性的、平面的、逐层展开的，也就是说文本内部的结构以由浅入深、由表及里，循序渐进的纵向联系为主，横向联系，如文中注释、引用、参见等为次。而超文本是网状的、立体的、联想式的，它在保留文本原有纵向联系的同时，为文中的知识点之间建立了大量的横向联系，这种横向联系可以是超越文本范围的，也可以是超越媒体的，并且是"点击式"立体的。所以，超文本从表面上看与传统文本无异，但却可以做到非线性、跳跃式阅读，文本中的任何一个知识点都可以成为阅读的视角，这些与其它知识有联系的知识点就称为节点，以高亮度文字或图标形式显示，知识点之间的联系就称为链路，链路建立在内部，是隐性的。

超文本突破了传统文本的线性单向阅读方式，它的最大特征是立体感与联想感，这两个特征都可以用来表现分类法的体系特点。

超文本的立体感可用来表现分类法的等级制，分类法的体系是层层划分、层层隶属的等级体系，每个上位类都有与其有概念隶属关系的一组下位类，建立链接，这样整个类目体系就成为立体的，类目层次感非常突出，传统信息环境下的类目加细，不仅会带来类表体积问题，而且用户检索起来颇为费力，而超文本技术的应用，符合了人们的探索欲望，类目深度的加深，不仅无累赘之感，反而有意犹未尽之感。而且，由于超文本中的上位类是包含一组下位类的一个节点，这样非常有利于多重列类法、双表列类法的使用，同时使用的几个分类标准不过是包含可以逐层展开的几屏信息、并列的几个节点而已。

超文本的联想感可用来表示分类法的类目横向关系。传统分类法,不管是纸本还是电子版,类目之间的横向关系是通过交替、参见或注释方式建立的,这些方式只能起到指引作用,用户还需到相应的类目重新查找,而超文本下交替、参见或注释或为一个与其它类目建立链接的节点,用户可以直接跳跃到所需的类目。这样,大量类目横向关系的建立,既不影响分类法的整个体系,又能使用户方便而迅速地浏览与检索相关信息。

分类法与超文本的利用是相辅相成的。超文本虽然提供了非线性、联想式阅读方式,但由于它提供的活动空间太大、选择的余地太多,人们在实际阅读中常常有漫无边际、无法把握的感觉,不是在浏览中迷失方向就是游离了原有的查询主题,忘记了原有的初衷,分类法则能有效地弥补超文本的这种缺陷。首先,分类法的聚类可以使用户的浏览限制在一个范围内,无论怎样从一个节点跳跃到另一个节点,都是与本范畴有关的主题;其次,分类法中建立在概念逻辑基础上的类目关系,可以控制和说明节点间的关系状态,避免节点间链接的盲目与无序,使用户心中有数地选择浏览视角。

网络环境下除了超文本技术可利用外,还有其它信息技术可利用,如利用多媒体技术来图形显示类目关系,利用人工智能来提高分类检索效率,利用网络技术来增强与用户的联系等等。在当今这个数字化生存的时代,技术的运用被提到了越来越重要的地位。

3.5 取长补短

取长补短包括两个含义,一个含义是指分类法借鉴其它情报检索语言的特长,如主题法、自然语言等,以弥补自身的不足。分类法在不影响整体特性与功能的前提下,可适当引进主题法等其它情报检索语言的方法,如在某些低层次类目采用字顺组织法,或设计字顺途径检索。推而广之,在整个系统的设计过程中,要考虑系统各个功能之间的互补互助,浏览工具与关键词搜索引擎的功能就是相互弥补的,搜索引擎提供的是自然语词检索途径,浏览工具提供的是分类查找途径,可以通过浏览工具引人受控语词,以提高检索效率。第二个含义是指检索之间的相互借鉴,不同的检索工具有不同的特点,基于分类法浏览工具与非基于分类法浏览工具之间就有各自不同的特长,研究与借鉴现有的网络信息分类组织工具的特点将有助于分类法的应用与发展。

4 结语

网络改变了分类法的生存与发展环境,改变了分类法的组织对象、处理方式、技术手段和服务用户,从而带来了分类法知识体系更新、类目体系改造、用户定位、技术利用等各方面的变化。但无论环境怎么变,新的信息环境下人们获取信息的需要和目的没有变,对于一种信息组织工具而言,它在网络环境下所要实现的基本目的和功能在本质上是和非网络环境一致的,只是实现的手段和环境不同罢了。同理,本文对分类法在网络环境下的发展提出种种建议和策略,最终可以归结到两个经典原则:文献保证原则和用户保证原则,这两个原则是本文的基本立足点。在一个新的、发生了变化的信息环境下,分类法要做的事就是了解环境的变化,包括信息中包含的知识的发展、用户需求与检索手段的变化,然后使自身逐渐适应这些变化。

另外要提的是,除了这两个原则外,分类法的电子化,被提到了越来越重要的地位,它是分类法在网络环境下生存必须具备的前提,网络环境本身就是得益于技术的进步,它也是现在和今后保持分类法生命力必不可少的手段。分类法的电子化,也不再仅仅是指类表的电子化,它包括多种信息技术的利用和跟踪、升级。

网络环境不仅促进分类法在实践方面的变革与发展,还将促进分类法理论框架的重构。文献

形式的丰富曾使图书分类学简单地更名为文献分类学,网络环境的到来却不能使文献分类学简单地更名为信息分类学,因为其中需要重新建设的东西太多,诸如分类法处理对象的再定义、分类法版本概念的再定义、分类法编制原则的再审思、分类法体系设计方法的再革命、分类法职能的再补充、分类法评价标准的再修订、分类法电子化内涵的再扩充、分类法修订方法的再补充、标引规则的再改革、标记符号类型的再总结、甚至分类学的再命名等等,分类法理论框架的重建还有待于深入研究与探讨。

参考文献:

[1]Traugott Koch et al. The role of classification schemes in Internet resource description and discovery: Specification for resource description methods Part 3.
http://www.ub.lu.se/desire/radar/reports/D3.2.3/class-v10.html

[2]董小英.网络环境下的信息资源管理(博士论文).北京大学信息管理系,1997

[3](印度)阮冈纳赞著,姚倩译.图书分类法原理 四-六章.图书馆界,1994(3):28-35

[4]赵文.面向最终用户:检索系统的趋势及其原则.图书情报工作,1996(4):55-57

[5]汪冰著.电子图书馆理论与实践.北京:北京图书馆出版社,1997

[6]王忠红.网络信息环境下的传统分类法.图书情报工作,1999(2):37-39

[7]Yahoo http://www.yahoo.com/

[8]Bubllink http://bubl.ac.uk/

[9]张琪玉.情报检索语言.武汉大学出版社,1984

[10]陈树年.基于计算机基础,对体系分类法改造的探讨.图书馆建设,1998(5):40-42

[11]王崇德,李美.论超文本信息系统.中国图书馆学报,1996(4):30-35

[12]David G. Dodd. Grass-Roots Cataloging and Classification: Food for Thought from World Wide Web Subject-Oriented Hiearchicdal Lists. Library Resources and Technical Services,1996,V,40(3):276-286

图书分类法性能的定量测评初探
——以《中图法》与《科图法》性能的定量测评为例

南京航空航天大学图书馆　尚加宁　韩露盈

[摘要]本文针对目前学术界对图书分类法的性能比较与测评一般以定性为主的情况，提出用定量测评的方法来更科学地评价图书分类法的性能表现，尝试确立了一些评价指标，包括了分类法类目数量及类目质量等方面的测评、类目交替度、参照度、清晰度等指标，并以《中图法》与《科图法》的性能定量测试为例，用上述性能指标较详细的数据定量测评了这两部当今中国应用最广的图书分类法。

[关键词]图书分类法　定量测评　性能指标

长期以来，《中国图书馆图书分类法》(以下简称《中图法》)与《中国科学院图书馆图书分类法》(以下简称《科图法》)以其良好的性能，简便实用的使用方法而在我国的图书馆界中被广泛采纳使用。这两部分类法的性能相比较而言，究竟孰优孰劣？多年以来，学者们大多对此做了不少定性描述与评价，但在对其定量分析和评价方面，却几乎是空白。由于定性分析难免会掺杂作者一定程度的主观随意性，所比较的差别又不会像泾渭分明般的，易造成"王婆卖瓜，自卖自夸"或者"公说公有理，婆说婆有理"这种众说纷纭的局面。定量分析用数字来说话，却往往比建立在定性分析基础上的评价更有说服力。本文尝试确定一些性能指标，拟采用定量测评的方法，更为客观地分析、评判这两部分类法，以得出明确的结论。

在此说明，鉴于《中图法》第四版及其配套索引尚未正式出版，我们采用两部分类法的上一个版本——中图法第三版与科图法第二版作为测评数据来源。

1　基于类目数量的评价

类目是文献分类法的基本单元，无论从科学性与实用性的统一，抑或宏观结构下微观结构方面对分类法的评价，必须建立在一定数量的类目集合基础上，类目数量是衡量分类法词汇完备性和专指度的一个重要指标，也是对分类法规模划分的依据，一般要求其类目划分详细、展开充分、数量众多，这样才能达到利于文献标引的目的。

我们先从整体上把握两部分类法的基本规模，请见表1，(见下页)。

从表1统计数据看，二法指导思想相同，均采用"五分法"作为其基本部分，各自派生出22与25个大类，均设有简表、详表、复分表，且其基本大类划分是相似的。科图法二、三级类目多于中图法，而中图法在主表及复分表类目总数上多于科图法。

表 1　总数据表

项目 \ 分类法	中图法(第 3 版)	资料法(第 3 版)	科图法(第 2 版)
基本部类	5	5	5
基本大类	22	22	25
二级类	214	214	296
三级类	1482	1488	1672
简表	1524	1521	436
主表	29421	48995	22493
复分表	1204	1322	757
类目总计	30625	50317	23250

以上我们从整体上比较了二表的规模，还有必要细分其主表部分。从社会科学、自然科学等主要部类中我们随机抽取了以下类目，希望找到二者在质量上的区别。

表 2　大类抽样数据统计表

	中图法			科图法		
	二级	三级	总数	二级	三级	总数
马列主义、毛泽东思想	7	38	119	9	40	125
社会科学——经济类	9	51	1419	5	32	1047
自然科学——数理化	5	53	1379	37	188	1059
技术科学——计算机科学	9	46	118	7	19	55
综合性图书	7	26	97	9	32	144
统计	37	214	3132	58	172	2880

从表 2 的各类数据比较中，我们可以看出：中图法类目总数仍大于科图法，在一些类目上，两者差不多，如马列主义、毛泽东思想。而在社会科学方面，一般是中图法篇幅较大。这种情况在自然科学的比较时正好相反，如表 2 所述数理化的类目数之比较，且从二、三级类目比较也可以得出科图法较重视科技领域文献的分类，这正是其特色与实际应用的体现。

复分表作为一种压缩篇幅、提高分面组配能力与标引能力的工具，有更广泛应用的趋势，类表数量也会随分类法的不断再版而有所增加。我们通过测评，得出表 3：

表 3　复分表数据表

		中图法	科图法
通用复分表类目	总论复分	56	50
	世界地区	315	277
	中国地区	49	43
	国际时代	14	19
	中国时代	30	98
	中国民族	58	63
	总计	522	550
专用复分表	专用复分表数	60	14
	类目总数	682	207

从表 3 可以看出，中图法在中国时代表、中国民族表两项比科图法类目少，而在专用复分表及世界地区表上则比科图法多。通用复分表总计中图法少于科图法 28 个类目，而在专用复分表数目

上则比科图法多46个表、475个类目。中图法正是通过多设置专用复分表的做法形成在利用复分表上的优势，弥补自己在一些学科领域上标引能力的不足。总体而言，二者在通用复分表上的差距是很小的，而且随着中图法第四版对中国时代表等的扩充，差距会更小。科图法第三版增加了世界民族表这一通用复分表。种种迹象均表明两表对复分表应用的重视是相似的。二者之间的差距主要还是体现在对专用复分表的使用上。

通过以上对两表主表及复分表内容的定量评价，我们认为中图法类目要比科图法丰富，并很好地使用了专用复分表，具有良好的适用性，因而能广泛流传使用。而科图法在保证一定规模类目的基础上，在其自然科学系统上分类更细，类目更多，体现了专业特色，是很符合专业图书馆、情报机构使用的，两表在其它领域均有良好的表现，难分伯仲。

2 基于类目质量的评价

在类目数量的测评的基础上，我们还必须选取一系列指标来测评其类目质量。类目质量能从根本上体现一部分类法在科学性与实用性上的结合程度，反映分类法具体的科学和实用性。正如一味追求规模效应的粗放型经济模式不能被现代经济所采纳一样，我们评价一部词表性能，当然也要“质”、“量”并举。我们一般采用以下几个指标来度量：

2.1 交替度

交替类目是为了处理具有多重等级关系的概念而在分类表中专设的类目。为使这类概念多向成类，在分类表有关位置同时设置该类目，可视实际处理需要而决定是否将其改为正式类目。这样，拓宽了分类表的适用性与生命力，处理一些交叉学科也更灵活方便。交替度是用来评价交替类目在分类表中所有类目所占的比率的指标，其计算公式为：

$$清晰度 = \frac{分类表内交替类目数之和}{分类表内类目数总和}$$

一般说来，交替度越大，分类表灵活性也越强，但如太大也易引起归类上的混乱。

下面，我们运用该指标，依旧选取表2所述5个类目统计其交替度，得结果如表4：

表4 交替度数据表

	中图法			科图法		
	交替	总数	交替度	交替	总数	交替度
马列主义、毛泽东思想	0	119	0.0%	26	125	20.8%
社会科学－经济类	48	1419	3.4%	43	1047	4.1%
自然科学－数理化	17	1379	1.2%	46	1509	3.0%
技术科学－计算机科学	1	118	0.8%	1	55	1.8%
综合性图书	4	97	4.1%	3	144	3.1%
统计	70	3132	2.2%	119	2880	4.1%

由表4可知虽然在某些类目上中图法交替度大于科图法，但在整体上，尤其是在自然科学领域科图法交替度要比中图法大，由此可见科图法十分善于利用交替类目的设置来增强该法的应用领

域,弥补在类目数量上的相对不足。

2.2 参照度

参照度是指分类法中每个类目接受参照项的平均数。设置参照项加强了类目间的联系,并有利族性检索及提高检全率与检准率,提供更多的检索入口词,从而提高检索效率。我们一般采用参照度这个指标来表示分类法在上述各方面的性能表现,其计算式为:

$$参照度 = \frac{参照项类目数}{分类法中类目总数}$$

下面,我们仍就上述5个类目所述,计算其参照度得表5。

表5 参照数据表

	中图法			科图法		
	参见	总数	参照度	参见	总数	参照度
马列主义、毛泽东思想	0	119	0	2	125	0.016
社会科学-经济类	10	1419	0.007	26	1047	0.025
自然科学-数理化	111	1379	0.080	43	1509	0.028
技术科学-计算机科学	7	118	0.059	4	55	0.073
综合性图书	0	97	0	1	144	0.007
统计	128	3132	0.041	76	2880	0.026

由表5可知,中图法在自然科学领域参照度大于科图法,而在社科领域这种状况刚好相反,这是由两个分类法各自特色所决定的。科图法有关自然科学类目设置得更多,划分得更细。相对而言,中图法更需借助相关参照来弥补自己在该方面的相对不足,这是由分类法的适用性能所决定的。

2.3 清晰度

清晰度较形象地来表达类目能被用户所理解并正确使用的可能性,也即分类法对类目的注释的详尽程度。注释一般有说明类目的使用方法、适用范围、类目定义及历史注释等类型。注释越详尽,其清晰度也就越高,越有助于用户明确类目的涵义及类目间的关系,越有助于用户掌握分类法的使用方法。当然考虑压缩分类法篇幅、降低成本等因素,清晰度也不能过高,过犹不及。清晰度的计算公式为:

$$清晰度 = \frac{注释数总和}{分类法中类目总数}$$

下面,我们运用该指标,选取二法对应类目,经测算,得表6。

表 6　清晰度数据表

	中图法			科图法		
	类目注释数	类目总数	清晰度	类目注释数	类目总数	清晰度
哲学	193	703	0.275	125	683	0.183
军事、军事学	133	379	0.404	185	420	0.440
自然科学(总论)	27	61	0.443	41	55	0.745
总计	333	1143	0.308	351	1158	0.303

由表 6 的结果我们可以得出中图法和科图法都善于运用注释来保证用户能正确使用分类法的结论，两者总计清晰度很接近这一事实表明，充分理解用户，想用户所想，是两法所贯彻之方针，也能更好地向我们解释为什么两法至今仍能长盛不衰地被广泛使用。

2.4　其它性能指标

其它测评图书分类法的性能指标还有分类法类目实际使用频率、标记符号性能、类目差错率及与分类法配套使用的分类法索引的质量与数量等。在此，我们不能一一展开了。

通过上述两部分类法在类目数量、质量方面的定量测评，我们可以认识到尽管中图法和科图法均采用了相同的指导思想，整体结构较类似，但中图法以资料法作补充，规模较大，许多性能指标表现均较良好，在类目总体数量(尤其社科类)、专用复分表、类目交替度与参照度等性能指标方面均胜于科图法。中图法的这种性能表现，使它在使用上更具广泛性与易用性，适用于大众意义上的、公众性质的一般图书馆、情报部门，而科图法尽管在类目规模、类目质量等方面的许多指标上稍逊于中图法，但它有专业上更优越的适用性，在自然科学领域分类精细、标识号码简单划一，使其作为最佳选择而被广泛应用于科学院系统、工农业技术类院校、厂矿企业等科研实力较强的部门。中图法与科图法在今后继续修订出版，完全有必要互相借鉴，共同发展，更广泛地为国内乃至世界范围内各类图书馆发挥其在学术与业务上的巨大作用。

参考文献：

[1]侯汉清，马张华主编．主题法导论．北京：北京大学出版社，1991
[2]侯汉清．索引法教程．南京：南京农业大学，1993
[3]侯汉清，王荣授主编．图书馆分类工作手册．北京：中国科学技术出版社，1992

主题词刍议

中国科学院文献情报中心　文榕生

[摘要]通过分析主题词及其相关概念的性质、内涵，探讨揭示文献主题的检索语言的类型及其特点，澄清了一些模糊认识；同时，探讨了主题词的发展趋势。

[关键词]主题词　检索语言　叙词　标题词　单元词　关键词

1　引言

主题词是检索语言的重要组成部分，是用规范的语词形式来揭示文献所具体论述、研究或表现的主要对象、问题或中心思想。然而，由于种种原因，汉语主题词在我国投入实际应用尚不多，有关概念的混乱更使不少人难识其庐山真面目。在当今，信息量呈几何级数增长，传统的检索方式显然难以适应，对主题词及应用进行认真探讨，不无必要。

2　叙词并不等于主题词

何谓“叙词”？在《现代汉语词典》、《辞海》等通用汉语辞书上难见其踪影，不少人对此则避而不谈，就是颇具权威性的诠释也语焉不详，甚至造成误导。试见以下数例：

——《汉语主题词表》认为：“主题词是以概念为基础的，包括正式主题词和非正式主题词，是经过规范化的并具有检索意义和组配性能的单词和词组，是用作标引和检索图书资料的依据。”[1]

——《文献主题标引规则》(国家标准·GB3860－83)认为：“主题词——又称叙词(即正式主题词)，它是经过规范化的词，在标引中用以表达文献的主题，在检索中用它构成提问式，以表达检索的需求。”“主题词表——又称叙词表。它是将文献标引人员或用户的自然语言转换成规范语言的一种术语控制工具；它是概括各门或某一学科领域并由语义相关、族性相关的术语组成的规范动态词典。”

——《汉语叙词表编制规则》(国家标准·GB13190－91)认为：叙词是“一种主题词，它是在文献标引与检索中用以表达文献主题而规范的词。”

——《中国大百科全书：图书馆学、情报学、档案学》中未设“主题词”条目，所列“主题词表”则见“叙词表”，它是“将文献作者、标引者和检索者使用的自然语言转换成规范化的叙词型主题检索语言的术语控制工具，亦称主题词表、检索词典。它是一种概括某一学科领域，以规范化的、受控的、动态性的叙词(主题词)为基本成分和以参照系统显示词间关系，用于标引存储和检索文献的词典。”[2]

——《文献叙词标引规则》(国家标准·GB3860－95)对《文献主题标引规则》修订后认为：“叙词：即正式叙词。标引与检索文献时，叙词表中规定用于表达各种概念的词。”

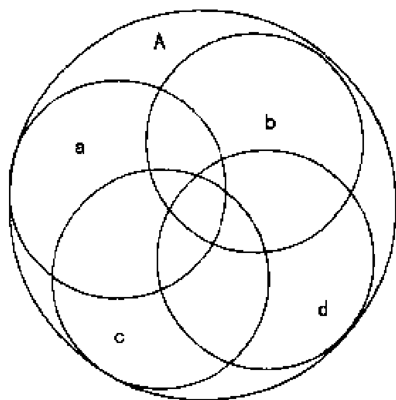
——《情报语言学基础》修订本则认为:"叙词(又称主题词)语言同标题词语言和单元词语言一样,是以自然语言词汇为基础的一种情报检索语言。"[3]

——经白国应、刘湘生先生指点迷津,笔者才进一步了解到:叙词是我国70年代引进的意译外来词,"叙词(Descriptor),也称描述词、叙述词。它是一些以概念为基础的、经过规范化的、具有组配性能、显示词间语义关系和动态性的词和词组。它是描述文献资料主题的一种标识符号。"[4]

尽管以上说法莫衷一是,但它们从各个角度反映出叙词的性质:叙述文献主题的概念单元,用语词形式揭示文献主题,对自然语言进行规范的语词,通过概念组配表达文献主题,受词表控制的规范词,显示词间语义关系和动态性的词,应用于检索系统。我们试将其归纳为:叙词是受词表控制,通过概念组配,叙述文献主题的语词型检索语言,它可以显示词间语义关系和动态性。

与叙词具有共性(规范的语词形式,反映文献的主要对象、问题或中心思想,应用于检索系统)的检索语言还有标题词、单元词和关键词,但它们又各有特性:标题词是采用预先搭配好的语词,直接反映文献主题;单元词脱胎于标题词,它是利用词汇的基本单元组配成文献主题;叙词综合了标题词和单元词之长处,采用布尔代数术语逻辑积、逻辑和与逻辑差作为文献主题和拟定检索命题(反映几个叙词之间概念组配关系的语法规则);关键词则直接抽取文献中表达主题概念的相关词。因此,这四种检索语言含有部分概念交叉,它们处于同一层次,彼此是并列关系。

而主题词只是反映文献主题的语词形式检索语言的概括,它与标题词、单元词、叙词、关键词是属概念与种概念的关系。如果用圈图(见文示图)表示:主题词是一个大圆(即图中A),标题词、单元词、叙词、关键词则是其中有部分互相交叉的4个相等的小圆(即图中的a、b、c、d)。



3 揭示文献主题的检索语言

检索语言是规范化的自然语言,它具有特定的涵义,是用于查检寻找有序化事物的工具。按照操作形式,检索语言可分为人工检索语言和机器(计算机)检索语言;再按表现形式,又可分为3种:专有型检索语言、代码型检索语言和语词型检索语言。而能够反映文献主题的检索语言除了语词型检索语言(即主题词)外,还有专有型检索语言(如题名形式)和代码型(如分类形式)。它们各有千秋。

3.1 题名形式检索语言与主题词的区别

从文字形式这一表象上看,题名形式检索语言(简称"题名")与主题词似相同,但二者实质上有区别:

3.1.1 揭示文献主题的表现形式不同。题名可以用词、短语,甚至句子来揭示文献主题;而主

题词多用词或词组来揭示文献主题。

3.1.2 揭示文献主题的程度不同。主题词直接反映文献主题;而学术性的,尤其是科技文献题名可以扣住文献主题,但也有相当部分题名(如文艺方面的)采用比喻、假借等手法并不能直接反映文献主题。

3.1.3 规范方式不同。题名可以直接在原形式上规范(如机读目录的200字段,5字段);而主题词则需要经过分析、判断、选择,采用恰当的,规范化的词或词组。

3.1.4 汇集文献不同。采用主题词可以汇集同主题(概念、涵义)文献;而采用题名只能汇集同字形或音的,但其主题却不一定一致的文献。

3.1.5 检索效果不同。采用题名检索,可以直接命中具体文献,查准率高;而采用主题词检索,尽管不一定直接命中具体文献,但查全效果较好。

3.2 分类形式检索语言与主题词的区别

分类形式检索语言(简称"分类")与主题词瑕瑜互见,它们的优势也可互补。

3.2.1 分类相对于主题词的特点。分类的稳定性、系统性明显,尤其是体系分类法系统性强,便于用户从知识体系(科学分类体系)查考;同时,将它应用于馆藏文献的排架、统计、咨询、宣传、研究等多项业务管理皆取得较好效果,这也是分类经历数千年不衰的主要原因。当然,分类的不足也是明显的:它抽象,不直观,必须使用辅助工具:它是先组式的,适应性有限,不能及时反映新事物;它的模式固定,积重难返,不易对其体系结构作重大修改等等。[5]

3.2.2 主题词相对于分类的特点。主题词固不宜用于文献排架等业务管理;但将它应用于检索,其灵活、直观、简洁的特点恰恰弥补了分类的不足。诚然,主题词是十分重要的检索语言,尤其适应现代科技综合、渗透、深入等纷繁复杂发展态势;但是,长久以来,主要受人员素质、工作量、技术、设备等制约,投入实际应用的单位寥寥。

4 主题词的发展趋势

现代科技日新月异,人员素质不断提高,设备不断更新换代,尤其是计算机的广泛应用,网络迅速发展,深深地影响主题词的发展。具体表现在:

(1)主题词的重要性加强。除了前述主题词在检索方面的优势外,叙词可以扩展应用到分类法的类目名称。[6]只要不同分类法的相同类目名称一致,便可打破目前多部分类法鼎立,又难以在短期内实现统一的僵局。

(2)注重名词的规范。随着知识经济扑面而来,信息交流更加活跃,新学科、新名词若潮水般涌现,使人目不暇接。例如:Internet的汉语名称曾出现"国际互联网"、"互联网"、"国际网"、"计算机国际互联网"等不下十余种,造成学术界和一般读者,特别是各种传媒体的名词用语严重混乱。经全国科学技术名词审定委员会建议;在科技书刊中宜直接引用英文原文,在必须用中文名词表达时则用"因特网"。通过规范,控制随心所欲增设新名词之风,便于文献标引与检索取得共识。

(3)叙词与关键词互补。《汉语主题词表》初编收录正式叙词91158条(其中社会科学领域23500余条,自然科学与工程技术方面67600余条),[1]增订后的自然科学与工程技术方面叙词为68823条;[7]而《科图法》第3版近4万个类目,《中图法》第4版据称约有5个类目,《国际十进分类法》设15-21万个类目,《美国国会图书馆分类法》设23万多类目。显然,目前的叙词量难以对文献进行适度标引。《科学技术期刊编排格式》(国家标准·GB/T3179-92)明确采用关键词,《文献叙

词标引规则》灵活地提出可采用“自由词”(应称“关键词”)标引，图书在版编目(CIP)数据中叙词与关键词共存现象不一而足。这两种主题词共存互补已被越来越多人士认同。

(4)编制电子版后控词表。计算机和网络的迅速发展、普及，为制作、推广、应用电子版后控词表营造了良好氛围。现有综合性叙词表又有近10年没有进行维护，加之各种专业性叙词表纷纷出台，将关键词与叙词有机联系起来等等状况，惟有通过后控词表，才便于语词的规范、标引、检索取得共识。

参考文献：

[1]中国科学技术情报研究所，北京图书馆主编. 汉语主题词表(第一卷，社会科学第一分册，主表：字顺表、试用本). 科学技术文献出版社，1980

[2]中国大百科全书总编辑委员会《本卷》编辑委员会，中国大百科全书出版社编辑部编. 中国大百科全书：图书馆学、情报学、档案学. 北京：中国大百科全书出版社，1993

[3]张琪玉. 情报语言学基础. 武汉：武汉大学出版社，1997

[4]刘湘生. 对叙词及有关几个名词术语的认识. 图书馆工作，1978(3)：27－28

[5]张玉麟，文榕生. 论文献检索语言的发展趋势. 图书馆，1995(4)：1－4，20

[6]文榕生. 文献分类法由被动走向理智：类目控制论. 中国图书馆学报，1997(6)：37－42

[7]中国科学技术情报研究所《汉语主题词表》自然科学部分维护组. 汉语主题词表(自然科学，第一分册，字顺表：A－L，增订本). 北京：科学技术文献出版社，1991

增词标引与自由词标引刍议
——兼议《文献叙词标引规则》的有关规定

福建省图书馆　张云瑾

[摘要]本文分析了增词标引与自由词标引的区别，提出增词标引应当严格控制，自由词标引应当放宽，《文献叙词标引规则》中有关自由词标引的规则应当修订。

[关键词]增词标引　自由词标引　文献叙词标引规则

1　前言

受控检索语言的基本原理在于词汇控制。词汇控制可以使词量缩小、词形统一、词义明确、词间关系清楚，从而提高主题标引的质量。然而，也正因为词汇控制，许多主题概念无法用用户熟知的专指词来标引，不得不进行组配标引、上位词标引或靠词标引，因此降低了标引的专指性，影响检索效率，造成用户检索困难。例如，在《中国分类主题词表》中表达“机读目录”这一主题概念的正式主题词是“机器可读目录”，而实际上“机读目录”这一语词更为人们熟知，在完全采用受控语言标引的检索系统中，用户如果按惯称输入“机读目录”这一检索词肯定检索不到与该主题有关的文献。类似情况数不胜数。

另外，主题词表的更新总是落后于事物的发展，新概念、新名词层出不穷，这些新概念、新名词采用词表无法标引。

为了解决上述矛盾，标引人员在实际工作中往往采用增词标引以及自由词标引的方法。这两种方法虽然都可以在一定程度上克服受控语言标引的局限性，但是两者存在很大的区别。然而，从近年来发表的有关情报检索语言和标引技术的论著来看，增词标引与自由词标引往往被混为一谈，由此在标引实践中出现了增词标引控制过松，自由词标引控制过严的现象，这对于提高标引一致性和专指性极为不利。为了避免再出现这种现象，笔者认为有必要对增词标引与自由词标引之间的区别作深入的研究。

2　增词标引与自由词标引的区别

增词标引是指当某个主题概念有较大的标引价值或检索价值，但是词表中未收这方面的专指词，而且又不宜采用组配标引、上位词标引或靠词标引时，采用经过规范化处理的新增加的主题词进行标引的方法。

自由词标引是指采用未经规范化处理的、词表以外的、具有较大标引价值或检索价值的自然语言词汇，如文献题名、文摘、正文中的关键词以及标引人员自己拟定的词汇对文献进行辅助标引的

方法。

具体来说，两者之间存在以下区别：

(1)采用增词标引的前提条件有两个：一、新增词必须具有较大标引价值或检索价值；二、词表中没有与新增词概念一致的专指词，而且也不宜采用组配标引、上位词标引或靠词标引。

采用自由词标引的前提条件只有一个：所要采用的自由词有较大的标引价值或检索价值。不必考虑词表中是否有与自由词概念一致的专指词，或者是否可以采用组配标引、上位词标引或靠词标引。

例如，"机读目录"这个词虽然具有较大的标引或检索价值，但是它不能作为新增词，只能作为自由词进行标引。

(2)新增词要进行规范化处理，即要在词量、词形、词义上进行控制，并建立相应的参照关系。

自由词基本上不必进行规范化处理，可以直接从文献题名、文摘、正文中选取，也可以由标引人员自己拟定。

(3)新增词一经采用，便成为词表中的正式主题词，采用增词标引之后，不必再同时采用组配标引、上位词标引或靠词标引。

自由词是自然语言词汇，在受控语言检索系统中只能作为辅助标引手段，在采用自由词标引的同时还必须采用词表中的正式主题词进行组配标引、上位词标引或靠词标引。

(4)在机读目录中，新增词著录在600～607正式主题词字段。起独立检索作用的自由词著录在610非控主题词字段，起修饰限定作用的自由词可以随被修饰限定的正式主题词著录在600～607字段。

3 严格控制增词标引，适当增多自由词标引

从增词标引与自由词标引存在的区别来看，要维护主题词表的权威性，保证全国标引数据的统一，便于资源共享，必须对增词标引进行严格控制。最好由词表编委会、国家图书馆等权威机构确定新增的主题词，并固定在某一刊物上发布，一般的图书情报部门根据权威机构发布的新增词进行增词标引。另外，也可以参考国家图书馆等国内几个具有代表性的编目中心发行的机读目录数据进行增词标引，即虽然词表中没有，但是几个编目中心的机读目录数据中使用的主题词原则上可以确定为新增词，用于本馆文献的标引。

与增词标引相比，应用自由词标引的条件可以适当放宽，标引人员可以根据用户的检索特点尽量标引出具有较大检索价值的自由词。这除了因为增词标引与自由词标引存在上述区别之外，还因为以下几个原因：

(1)由受控语言向自然语言方向发展是情报检索语言发展的必然趋势。随着国际互联网的发展，情报检索的操作者逐渐由受过专业情报检索技能训练的图书情报工作者转变为学科不同、水平参差不齐的一般用户，要他们熟练掌握受控语言检索技术是很困难的。因此，他们对情报检索语言自然语言化的呼声越来越高，满足他们的情报检索需求是图书情报工作者的责任，但是由于技术条件的限制，完全采用自然语言标引和检索，目前还达不到理想的检索效率，只有将受控语言和自然语言相结合才能实现高效率的检索，而以受控语言标引为主、以自由词标引为辅正是目前最切实有效的结合方式。

(2)放宽自由词标引是弥补我国主题词表本身缺陷的必要措施。据统计，《分面叙词表》中叙词的等同率(即非正式叙词与正式叙词的比率)为43.75%，《Unesco叙词表》为53.71%，《基础叙词

表》为46.61%,《教育主题词表》为22.95%,《中国分类主题词表》为19.10%。等同率越低就意味着词表提供的检索入口词越少。从以上统计数字来看,我国叙词表的检索入口词大大少于国外叙词表,国外几乎达到正式叙词的1/2,而我国只有1/5左右。众所周知,检索入口词越多越有利于提高检索效率。为了弥补我国词表的先天不足,标引人员应当尽量采用自由词进行辅助标引,以便为用户提供更多的检索入口。

(3)计算机在图书情报工作中普遍应用使放宽自由词标引成为可能。在手工检索系统中,每增加一个检索入口就至少要增加一张款目,如果大量增加自由词标引则无疑会大大增加图书情报部门人力、物力和财力的投入,因此不得不对自由词标引进行严格控制。在机检系统中,每增加一个检索入口只需增加一个字段即可,放宽自由词标引不会增加太多人力、物力和财力的投入。另外,在机检系统中,CNMARC格式将正式主题词与自由词区别对待,分别采用不同的字段著录,这样可以使自由词标引检全率低的缺陷得到弥补,因为用户可以根据检索词在书目记录中所处的字段来判断所输入的检索词是正式主题词还是自由词,如果是自由词,则再采用同一记录中正式主题词字段的有关主题词进行检索,这样就可以大大提高检全率;如果采用设计合理的软件,计算机也可以根据检索词所处的字段来自动提示某个检索词是自由词还是正式主题词,从而引导用户再从正式主题词入手进行检索,避免漏检。

4 《文献叙词标引规则》中有关自由词标引规则的商榷

从以上分析来看,笔者认为《文献叙词标引规则》(下称《规则》)第4.5条有关自由词标引的规则已不能适应文献标引和检索的需要,应当重新修订,具体修订内容如下:

“4.5 如果不宜用4.4以前的各种方法进行标引时,可以采用增加经过规范化处理的新的叙词的方法进行标引。增词标引应当严格控制。

4.6 自由词是未经规范化处理的叙词表以外的自然语言词。在采用叙词表中的正式叙词进行标引的同时,可以根据需要采用有较大标引价值和检索价值的自由词进行辅助标引。

4.6.1 下述情况可采用自由词进行辅助标引:……”

余下条款与《规则》基本一致,各条款的编号应当作相应调整。

参考文献:

[1]陈树年.中国分类主题词表标引手册.北京:北京图书馆出版社,1998
[2]侯汉清.索引技术和索引标准.北京:北京图书馆出版社,1997
[3]司莉.五部分类主题一体化词表的比较研究.图书馆,1998(5):32-35
[4]侯汉清,马张华.主题法导论.北京:北京大学出版社,1991

我国自然语言检索研究的历史、现状及趋势

南开大学信息资源管理系　李新华

[摘要]我国自然语言检索研究走的是“自然语言－受控语言－自然语言”的发展道路。自然语言检索研究呈现如下特点：自然语言与受控语言结合探讨为主线贯穿整个研究过程，一定时间段内又有研究重点。笔者依据研究重点把我国自然语言检索研究划分为三个阶段：自然语言标引与自动分词研究阶段(1982－1988)；单汉字标引与检索研究阶段(1989－1993)；后控词表研究阶段(1994－至今)。然后分别讨论了这三个阶段的发展情况，最后预测了我国自然语言检索研究的发展趋势。

[关键词]自然语言检索　自动分词　单汉字检索　自然语言标引　后控词表

1　序言

自然语言检索，从技术上讲就是将自然语言处理技术应用于情报检索系统的信息组织。标引与输出，从用户来说就是用自然语言作为提问输入和对话接口的检索方式。[1]在建立情报检索系统之初，自然语言作为检索语言工具，受到学者们的广泛关注。随着研究的深入，自然语言在手工检索系统中越来越体现出其缺点，作为对自然语言进行控制的手段—受控语言诞生、发展并成熟起来，自然语言受到冷落。但随着检索理论和技术的发展及计算机在情报检索中的运用，受控语言缺点暴露，曾一度被受控语言否定的自然语言在检索中的地位又重新得到肯定。[2]这种肯定是在对原有自然语言否定的基础上的再次否定，是更高层次的肯定。可以这样说，自然语言检索研究符合事物发展的“螺旋式”上升规律。

2　我国自然语言检索研究发展概述

国外，对自然语言检索的研究起步较早，大约是50年代开始，如今在自然语言接口等与情报检索语言结合使用方面取得很大成绩。相比之下，我国起步较晚。我国自然语言检索的研究是随着情报检索计算机化发展起来的。开始，有学者介绍国外借助于计算机的自然语言检索法，随后，我国就开展了自然语言检索研究。我国最早以论文形式公开发表的对自然语言进行研究的是陈光祚老师。他于1982年《索引语言及其逻辑》一文中认为：自然语言属于句法语言(索引语言包括句法语言和非句法语言)，索引语言应当接近自然语言。[3]可以这样说，这篇文章首开我国自然语言检索研究之先河。后来在80年代，我国掀起一股研究自然语言检索的热潮，大量文章发表，还有几本专著问世，尤以张琪玉老师的《情报语言学基础》为代表。进入90年代，研究的少了。

笔者观察到，在我国自然语言检索研究过程中，呈现这样一个特点：自然语言与受控语言的对比分析、结合研究为主线贯穿整个研究过程，一段时间内又以某一方面为研究重点。就拿自然语言

与受控语言结合问题来说，从1982年到目前，历年都有文章进行探讨。最早是王知津老师在《受控标引与非控标引评述》一文中，提出受控语言与非控语言结合。[4]接着，有学者探讨二者的结合方式：规范词表的入口词汇；两种语言词汇同时标引一篇文献，构成同一词档法；标引用自然语言，查找用后控词表控制查找词法。[5]这是最早对二者如何结合进行探讨的文章。有学者认为，在联机检索中，可将自然语言与受控语言结合，以受控语言为主以自然语言为辅。[6]还有学者认为二者可以通过事先控制法、事后控制法、混合词表来兼容。[7]后来又有学者建议将自然语言与受控语言结合成整体化检索语言进行标引检索。[8]不同于以上的一种结合方式是：受控词表的自然语言入口词；专用入口词表；同时使用自由词与受控词；自然语言标引，后控词表检索。[9]这些结合方式都是从理论上进行探讨，没有试验证明其可行性。但是，对实际运用却有一定指导意义。自然语言与受控语言的结合方式、方法、模式研究还有待于我们进一步深入。

就阶段性研究重点来说，笔者注意到，从我国正式开始研究自然语言检索到目前，呈现三个研究重点的依次发展：1982－1988，探讨自然语言标引与自动分词并提出一种模式的文章居多；1989－1993，单汉字标引与检索研究成为重点；1994－如今，又以后控词表研究为主。当然，这并非严格的界限划分，只不过这一阶段这方面研究比较突出。因而，笔者姑且以这三个研究重点来划分我国自然语言检索研究阶段，划分如下：

1982－1988，自然语言标引与自动分词研究阶段；

1989－1993，单汉字标引与检索研究阶段；

1994－如今，后控词表研究阶段。

下面就对这三个阶段进行分析。

2.1 自然语言标引与自动分词研究阶段(1982－1988)

自然语言标引研究在国外很早就有人进行探讨。著名的克兰菲尔德试验就从实践上确认了自然语言标引的优势：自然语言在消除其同义词和将字形有变化的词进行混合的情况下，所得结果较其它索引语言更为优越。[10]自然语言标引可以是人工的，可以是自动的，也可以是不标引。目前主要是建立在自动抽取文献篇名、文摘中的关键词基础上，人工赋词的自然语言标引没有优越性。在我国，首次明确提出用自然语言进行标引的是赵宗仁。他于1984年的全国第四次机检会议上提出，采用自然语言标引具有很多优势，应该采用自然语言标引，[11]在当时引起不小反响。此后有学者把采用有控检索语言的赋词标引和采用自然语言的抽词标引(自由词标引)进行对比分析，认为采用自然语言作为检索语言能从根本上解决检索语言的专指度问题。[12]还有学者对自然语言标引的现状及趋势做了概括。[13]自然语言标引与自然语言检索联系密切，自然语言标引问题解决好了，检索就容易得多。今后要注意加强自然语言标引的深入研究。

在情报检索系统中应用自然语言是以语言自动处理(自动抽词)为核心的，即首先要从文献原文中抽出关键词，然后再对所抽出的词进行其它处理。这种抽词在国外一般是指关键词标引，在国内，由于汉字的构词与书写和西文不同，自动抽词十分困难，于是便出现了自动分词研究。从开始研究自然语言检索起，国内大部分研究就集中在自动分词模式的探讨上。

1983年，陈培久老师在他的汉语科技文献标题的自动标引试验中，提出词典切分组词法。他利用含有语法和语法加工信息的机器词典对文献标题进行切分，并将切分结果用40种规则组词。[14]

1984年，王永成老师进行了汉语部件词词典切分抽词的研究，[15]基本原理是：建立一字或二字汉语部件词词典，逆向扫描，二字前进，从部件词词典中抽出部件词，并根据词频和语言学分析得

出的 16 种词性进行组配标引。后来,他又进行了补充修改[16-17],进一步增加了部件词词典切分的实用性。

1986 年,吴蔚天在第五次全国机检会议上提出非用字后缀表分词法。[18]该方法是将汉字分为"非用字"、"条件用字"、"表内用字"、"表外用字"4 类,根据词表将标题或句子切分成短语和词组,再参照 6 种切分模式分解成单词或专用词供标引之用。

1987 年,邓钦和、龙泽云提出统计分析法进行切分。[19]这种方法采用逆向扫描、最长匹配的原则分词。是停用词,则放弃并逆向取下一段,关键词记下,否则作为备选词登记。当备选词达到一定频率时,转为关键词。选出的关键词必须进行词频统计和位置加权处理确定标引词。词表可通过备选词的频率统计从无到有地自动建立和扩充。

1987 年,北京大学图书馆系的主题词表法,[20]抽词手段由"主题词表"、"禁用词表"及逻辑判断规则构成。抽词方法:用"禁用词表"排除无用词,剩下短语与"主题词表"匹配抽出结果词,再利用汉语局部语法规则和一些主题判断规则,确定出主题词。

后来,许多学者陆续提出汉语词素自动分词法、"语境相关"自动分词方法、链接表法、语词结果类比法等。这些分词模式立足于机械分词,因此,后来有学者提出语法语义分析分词法、神经网络分词法等一些先进的分词方法。由此可见,我国的自动分词技术取得了一定进展。归根结底,以上分词模式可归为以下三类:

(1)词典分词法,代表为词典切分组词法、部件词典法、主题词表法等,这些方法的原理基本相似,都有两个或多个一定量的词表和一些规则组成。

(2)切分标记分词法,如非用字后缀表法、链接表法等,它们都是通过用切分标记将文本割成词组或短语,然后再按一定的分解模式分割成单词或专用词。

(3)其它分词法,如统计分析法、语法语义分析法、语法结构类比法等。[21]

十几年来,我国在自然语言检索方面研究最多的最成熟的就是自动分词技术。由上面可以看出,我国自动分词问题的解决指日可待了。但是,这个问题的解决,只是达到了拼写文字国家的起点水平,拼写文字中的问题还很多。

2.2 单汉字标引与检索研究阶段(1989－1993)

由于汉语分词的困难,一些学者独辟蹊径,仿照西文关键词法,提出单汉字标引与检索模式,即将文本中全部汉字(也可排除无用字)由计算机自动做成单汉字倒排档,检索时用单汉字组配法查找的一种检索方式。

据笔者所知,最早用计算机建立单汉字检索功能工具的是武汉大学计算机科学系与中文系联合的语言自动处理研究组。他们从 1978 年开始利用计算机对汉语书面语言进行自动处理,为《骆驼祥子》等文学作品自动建立"逐字索引"。其功能是自动统计不同汉字的出现频率,提供各单汉字在文本中出现的页号、行号,并可检索含有任意汉字的句子。[22]

80 年代,广州市科技情报研究所对小型的书目数据库中的题名字段进行了单汉字检索,并发表《单汉字机助索引和检索》一文。在文章中,他们提出了建立单汉字检索系统的基本思想和实现方式,并分析了单汉字检索的优点。[23]这是我国最早公开提出此模式的文章,同年,有学者在科技文献计算机标引中也提出单汉字标引。[24]

接下来几年,一股研究单汉字的热潮兴起。

1990 年,经济日报社与北方交通大学研制成功"经济日报新闻资料检索系统",并通过了鉴定,这是针对大型全文数据库实现的单汉字检索模式。这个系统已投入使用,从而从实践上证明了单

汉字检索模式的可行性。[25] 同时，苏新宁老师对单汉字标引算法进行改进，引入"位置运算"概念。[26-27]

1992 年，陈光祚老师全面系统论述了单汉字检索模式的历史发展、可行性及优缺点，[28]这篇文章囊括了单汉字的几乎所有内容。就在这一年，有学者尝试单汉字全标引检索系统。[29]

1993 年，许多学者讨论单汉字的各个方面，包括：系统检索效率的保障[30]，中文题录单汉字和关键词混合检索系统研究[31]等。还有学者将单汉字检索系统发展归纳为三代[32]：第一代是以无条件单字标引，单字逻辑与运算为特征的纯单汉字检索系统，如参考文献[23]；第二代则引入了西文全文检索中的"位置运算"概念，从而使查准率提高，如文献[26－27]；第三代是 90 年代出现的，采用非用词典控制索引数量，并以后控机读词典辅助检索，如文献[28－29]。这种划分比较准确。后来，针对单汉字标引检索算法陆续有学者进行改进。

单汉字标引与检索是符合我国汉字特点的检索模式，它逐步在理论和应用研究上得到发展完善。它的优点显而易见：避免了分词的障碍，能及时处理新概念词，组配灵活等等。但是，单汉字检索以单字为处理单位，又容易产生虚假组配，检索噪声大而筛选负担重，检索效率较低，扩检、缩检、改检比较困难，所以有学者提议用采用后控词表来弥补这些缺陷。

2.3 后控词表研究阶段(1994－如今)

后控词表是受控语言与自然语言结合的一种方式，它是将同义词与相关词搜集起来而编成的一种只供检索不做标引之用的词表。在我国，最初称"只检索词表"。这个单词首次出现在王知津老师《受控标引与非控标引评述》一文中。最早把后控词表上升到理论研究的是臧国权。他在 80 年代中后期连续发表两篇文章对后控词进行探讨。[33-34]在文章中，他分析了后控词表的功能、结构，阐述了后控词表编制方法。

后控词表的研究高潮是在 1994 年，这一年论文数量较多，研究内容包括：后控词表在情报检索系统中的作用[35]，后控词针对自然语言是一种控制措施[36]，后控词表的编制及设计方法[37-38]。其中研究最全面的数张琪玉老师的《论后控词表》一文[39]，他的文章涉及后控词表的控制机理、控制程度、编制特点、编制方法等。

1995 年，赖茂生老师等人对基于超文本结构的后控词表管理系统进行了初步开发模拟，并研制了一个模型。[40]

1996 年，周全明在其硕士毕业论文中就全文检索系统后控词表的后控制技术、后控词采集政策、后控词表范畴体系确立方法等方面进行了探讨，并设计出一个具体的全文检索系统后控词表。这篇论文分四部分公开发表了。[41-44]

随后，陆续有学者研究后控词表的其它作用及其结构，后控词表研究逐步成熟。因而，有学者建议在单汉字检索系统中配备后控词表，这无疑在很大程度上提高了系统的检索效率。[45]随后还有学者对单汉字检索系统中的后控词表进行了理论和技术上的改进。[46]

后控词表综合了自然语言和受控语言的几乎所有优点，成为目前自然语言和受控语言结合的典范。所以，后控词表前景十分广阔。

3 我国自然语言检索研究发展趋势

我国自然语言检索研究发展到如今，在理论和技术都取得了一定进展，如自动分词技术日益完善，一些自然语言检索系统初步试验应用，但是，距离国外差别还很大。自然语言检索研究范围广，

内容深,我国的研究仅限于一两个方面,大部分课题有的才开始,有的还是空白,这不能不说是一个遗憾。因此,我们要加强自然语言检索的全方位研究。

在情报检索系统中,自然语言的优势越来越显著,但这并不意味着自然语言将会取代受控语言,毕竟受控语言的某些优点是自然语言无法替代的。不论今后自然语言检索系统如何发展,都不可能取消控制,受控语言的痕迹将会永远存在。随着我国情报检索计算机化和自然语言化的进一步发展,自然语言和受控语言结合的趋势将更加明显。因此,对它们的结合方式、方法及模式的研究成为今后很长一段时间内的研究重点。

自然语言检索有着其它语言无法相比的优势,但也明显存在一些缺陷。由于自然语言的不规范,存在大量同义现象、多义现象和模糊现象,概念的隐含含义显示不出来,因而检全率低。所以,我们要探讨各种措施,来减少自然语言的缺点,充分发挥自然语言检索的优越性。

后控词表的编制情况很大程度上决定了自然语言检索的效率,后控词表的出现就是为了弥补纯自然语言检索系统的"不受控制"性产生的缺陷,所以,加强后控词表的研究极为重要。

其次,以自然语言应用为核心,研究其应用条件、检索效率、成本效益等现实问题,尤其是在网络环境中的应用情况;并从对某种具体语言的结构、功能、特点转向更为抽象的控制手段、控制思想的研究。

还要注重入口词表、规范词典和其它有关规范语言的自然语言接口问题的研究。

探讨各种情报检索语言的原理、方法在自然语言中移植的可能性及实现途径,弥补自然语言检索缺点,以便更好地完善自然语言检索。

情报检索语言的自然语言接口及人工智能的自然语言理解研究还是空白,没有引起足够的重视,今后这一研究将成为自然语言检索研究的热点。

参考文献:

[1]李法勇.谈自然语言检索的发展.情报理论与实践,1997,20(5):260-262
[2]戴维民.论我国情报检索语言研究道路.情报学报,1988,7(2):114-121
[3]陈光祚.索引语言及其逻辑.图书情报知识,1982(4):15-22
[5]汪东波.规范语言与自然语言情报检索.图书情报知识,1986(3):33-34
[6]钟家瑞.论文献检索的受控语言与自然语言.情报业务研究,1988,5(1):29-37
[7]徐燕萍.论受控语言与自然语言的兼容.情报业务研究,1988,5(1):24-28
[8]洪漪.我国信息网络建设中的检索语言问题.中国图书馆学报,1995(3):69-71
[9]戴维民.自然语言标引与检索的发展趋势.情报业务研究,1991,8(4):252-254
[10]陈光祚.各种索引语言与标引系统的比较分析:著名的"克兰菲尔德试验"概述.图书情报知识,1983(1):12-18
[11]赵宗仁.论自然语言标引问题.全国第四次机检会议论文,1984.又见:情报学报,1985,4(1)42-
[12]夏国梁等.论检索语言在标引中的作用和地位.情报业务研究,1985(3):135
[13]同[9]
[14]陈培久.汉语科技文献标题的自动标引试验.情报学报,1983,2(2):113-120
[15]王永成,肖玮英.自动编制中文标题的主题词轮排索引及自动抽词.南京大学学报:自然版,1984,20(1):39-44
[16]王永成等.自动编制主题词合成部件词典.南京大学学报:自然版,1985,21(1):93
[17]王永成.论从中文资料中自动抽取关键词.图书与情报工作,1986(2):36-38
[18]吴蔚天.在微机上实现汉字抽词.全国第五次机检会议论文.西安,1986.5
[19]龙泽云、邓钦和.一种汉语自动分词标引方法——统计分析法.现代图书情报技术,1987(3):11-13

[20]北京大学图书馆系.汉语科技文献自动标引系统.情报学报,1987,6(4):277-284
[21]张琪玉著.情报语言学基础(增订2版).武汉:武汉大学出版社,1997,9
[22]陈广祚.论单汉字检索系统.情报学报,1992,11(1):11-18
[23]黎小林,吴骏盛.单汉字机助标引和检索.情报学报,1988,7(1):29-33
[24]高崇谦.我国科技文献计算机标引研究的现状.情报学报,1988,7(1):67
[25]同[22]
[26]苏新宁.中文单字标引算法的改进设想.现代图书情报技术,1989(1):14-16
[27]苏新宁等.论中文标题的单字标引与位置检索.南京大学学报:自然版,1990,26(2):329-335
[28]同[22]
[29]刘继昌.全标引汉字情报检索系统研究.情报科学,1992(1):19-22
[30]汪红秋.单汉字检索系统查全率查准率的保障.情报科学,1993,15(6):441-445
[31]夏海.中文题录单汉字与关键词混合检索系统.现代图书情报技术,1993(1):22-23,41
[32]夏景峰,李必旺.面向自然语言查询的单汉字检索系统.情报学报,1995,14(6):410-416
[33]臧国权.后控词表浅谈.图书与情报工作,1986(4):74-78
[34]臧国权.后控词表检索系统.情报科学,1988(2):42-50
[35]陈道泉.情报检索系统与后控词表.现代图书情报技术,1994(4):48-50,5
[36]宋明亮.论对自然语言的控制.情报理论与实践,1994(2):27-30
[37]陆长旭.后控词表的编制方法.中国图书馆学报,1994(6):18-21
[38]韩冬梅.后控词表的设计开发与利用.情报学报,1994,13(4):266-272
[39]张琪玉.论后控制词表.图书情报工作,1994(1):1-5
[40]赖茂生,谭晓冬.基于超文本结构的后控词表管理系统.情报学报,1995,14(5):321-328
[41]周全明.关于全文检索系统后控制技术研究的基本观点和结论.图书馆杂志,1996(4):19-21
[42]周全明.全文检索后控关键词采集政策研究.情报理论与实践,1996,19(4):35-39
[43]周全明.全文检索后控词表体系确立方法初探.情报理论与实践,1996,19(6):14-16
[44]周全明.一个具体的全文检索后控检索系统的设计.现代图书情报技术,1996(3):32-36
[45]同[22]
[46]丁蔚.单汉字检索系统后控词表的改进研究.现代图书情报技术,1998,15(5):25-28

积极为自然语言与情报检索语言的结合创造条件
——建议大量编制自然语言词表

空军政治学院信息管理系　张琪玉

[摘要]本文从文献标引与检索的实际情况出发,概述了当前需要的自然语言词表的类型,着重阐述了自然语言词表编制中的方法和技术问题。

[关键词]自然语言接口　自然语言检索　自然语言词表　编制技术

1　大量编制自然语言词表在促进文献标引－检索用语言进步中的作用

在本文中,"自然语言词表"一词是指含有自然语言成分的各种词表,或者说自然语言应用于情报检索所需的各种词表。

1.1　文献标引－检索用语言发展的大趋势

文献标引－检索用语言目前有情报检索语言和自然语言。情报检索语言的未来和自然语言的未来,即两者发展的大趋势,是情报检索语言的自然语言化和自然语言的情报检索语言化,是它们从互相结合到完全融合的过程。未来的文献标引－检索用语言决不会是纯粹的自然语言,即自然语言不可能取代情报检索语言;也决不会是纯粹的情报检索语言,即不会保持目前的这种人工语言模式。总之,情报检索过程绝对不能没有控制,情报检索语言的控制原理将依然被保存,但将来的控制模式不再会是现在的控制模式。或许,未来的文献标引－检索用语言将仍然称为情报检索语言,因为它们对情报检索过程进行控制的本质不会改变。

1.2　文献标引－检索用语言当前的情况

文献标引－检索用语言当前的情况是:

(1)情报检索语言当前已达到相当高的控制水平,进一步的改进主要是在易用化方面。学者们几乎一致认为,其主要的易用化措施是采用与自然语言相结合的各种方法,或者说增加自然语言成分,也就是情报检索语言的自然语言化改造。

(2)自然语言在情报检索中的应用,当前正在迅速发展。但自然语言检索法尚处于其发展的初级阶段,它既有许多显著的优点,也有许多突出的缺点。以往有许多著作认为检准率高是它的显著优点,但从文本检索的实践中暴露出这一优点也不是普遍存在的,有时检准率甚至低到不能容忍的地步。从情报检索过程绝对不能没有控制这个基本原理看,它的各种缺点的根源在于缺乏控制,所以它的进一步改进还是要引进控制原理,即情报检索语言所采用的各种方法的原理,也就是自然语言的情报检索语言化改造。

(3)当前,网上知识－信息资源的蕴藏数量和增长速度是惊人的,这些资源是取之不尽、用之不竭的财富。网上知识－信息资源大多未用情报检索语言作标引,所以自然语言是其主要的检索用语言,现有的网上检索工具——搜索引擎也亟待改进。

1.3 自然语言词表在文献标引－检索用语言进步中的关键作用

自然语言词表是自然语言与情报检索语言结合的必要条件,可以预见,它在促进当代文献标引－检索用语言的进步中将起极大的关键性的作用。因为:

(1)自然语言词表是提高情报检索语言易用化的主要手段。情报检索语言的自然语言化,无论采用哪种方式,都离不开某种类型的自然语言词表。有自然语言词表辅助的情报检索语言标引和检索将是一个重大的进步。

(2)自然语言词表是对自然语言加以控制的主要形式。因为自然语言词表就其实质而言,是对庞大纷繁的自然语言词汇进行某种程度的系统化整理,采用或多或少的控制方法。在自然语言词表辅助下的自然语言检索,相对于无自然语言词表辅助的自然语言检索来说,检索效率必然会有较大提高。所以,有自然语言词表控制的自然语言检索也将是一个重大的进步。

2 当前需要的自然语言词表的类型

2.1 自然语言接口用对应表

自然语言接口可以说是在情报检索语言的自然语言化中最简易可行的一种方式。它实际上是在情报检索语言检索系统之前安置一个自然语言语词与情报检索语言语词的对应表,其前端(被转换的源词字段)为自然语言的语词,后端(所转换成的目标词字段)为情报检索语言的语词。检索人员(或标引人员)使用自然语言的语词表达检索课题(或文献主题)进入系统,通过对应表自动转换成情报检索语言的语词在系统中进行实际的检索(或标引)。对应表的内容可以逐步积累扩充,任何规模都可投入实际使用。对应表只是一个附加部分,并不影响原有的标引工具和标引数据,所以,使用自然语言接口只会增益,是有利无弊的,可普遍采用。

自然语言接口的对应表其实就是入口词表的机读版,可比书本式入口词表的收词范围更为宽松。在对应表中,自然语言与情报检索语言的对应可以有一对多的关系,通过人工辅助转换。为使对应表一目了然以便于管理,并简化转换,可将词表正式词也作为一个自然语言词重复列入对应表,作成“词表正式词→词表正式词”对应款目。词表的双语种对照索引也可编入对应表。

分类表索引的机读版与对应表本质相同,可作为自然语言接口的对应表使用,如果能再增加大量自然语言语词,则更好。

2.2 自动抽词词典

在汉语自然语言检索中,抽词标引系统比无标引系统(包括仅提供任意字词匹配检索即模糊检索功能的系统和单汉字系统)好。所以,汉语自动抽词技术应成为汉语自然语言检索的主流技术。汉语的自动抽词系统,绝大多数都是使用抽词词典的。抽词词典不但是抽词标引系统实际投入使用所不可缺少的条件,而且对抽词质量还具有重大影响。

近 20 年来,我国学者对汉语分词(即汉语自动抽词)技术作了许多研究,提出了不少分词方案、通过了鉴定,但见于实际使用者不多。并不是这些方案都经不起实践考验,而主要是因为半途而

废。因为只有抽词软件而无抽词词典，是不能建立自动抽词标引系统的。编制了抽词软件而不编制抽词词典，事情只能说做完了一半，甚至只是完成了很少的一部分，因为编制抽词词典比编制抽词软件可能需要许多倍的工作量。目前缺乏抽词词典，已成为自动抽词标引技术（此项技术也是自动赋词、自动赋号、自动分类等的基础）难以普及的主要原因。

当前迫切需要大量编制汉语自动抽词词典。之所以“需要大量编制”，是因为专业性的抽词词典较易编得好，而且专业范围越小越易编得好，故需要上百部甚至更多的抽词词典，才能满足需要。

我国也需要外语的自动抽词词典，即非关键词表。非关键词表也称索引不用词表、停用词表或禁用词表。

2.3 自动赋词赋号用对应表

自动赋词赋号标引系统是对自动抽词标引系统的改进，使自动抽出的自然语言语词转换成情报检索语言语词（检索词或分类号）。其所用的对应表的前端（被转换的源词字段）为自动抽词所抽出的自然语言语词，后端（所转换成的目标词字段）为情报检索语言语词（检索词或分类号）。自动赋词标引系统和自动赋号标引系统可以分别建立，也可以合而为一。合而为一时，对应表设两个目标字段，第一个目标字段为检索词，第二个目标字段为分类号。其实，设立更多个目标词字段也是可以的。

自动赋词赋号标引系统可使用现有的词表或分面分类表，也可仿照词表和分类表的编制原理，对自动抽词所抽出的自然语言语词作有限范围的控制（如只对同义词作规范控制，或纳入粗略的分类体系）。

自动赋词赋号标引系统还可以在赋予文献检索词或分类号的同时，仍保留自动抽词过程所抽出的原词，兼取人工语言与自然语言的优点。这样的系统用于检索时，检索者既可使用检索词或分类号检索，也可使用自然语言检索。应当指出，使用现有词表的自动赋词系统和使用现有分面分类表的自动赋号系统，从检索角度看，也都是一种自然语言接口。

2.4 自动分类用对应表

2.3 节所述自动赋号标引其实也是一种自动分类。本节所述的自动分类与其不同处在于：它使用体系分类法，自动分类得到的分类号可区分出主要分类号和非主要分类号，各个分类号的组配又可表达比原有类目更多和更专指的概念。其所用词表是一种词与分类号的双向对应表，由分类号－词对应表和词－分类号对应表两部分组成。

分类号－词对应表的编制法如下：假定使用《中图法》，需先将《中图法》的分类表改造成分面分类表，但原有分类号不需要改变。把自然语言语词对应到相应的分面中。由于文献主题一般都是由多个主题因素构成的，各个主题因素在体系分类表中都有其对应的类目。例如，“客车传动系统设计”这个主题，分入“U463.1 汽车传动系统”、“U462 汽车设计”、“U469.1 客车”三个类目或其中的任何一个，从事物概念的多向成族性看，都是正确的，以哪个类目作为主要类目，这只是“集中与分散”的需要不同而已。依据一般分类规则把上述主题分入”U469.1 客车”类目，就是按事物集中，是把该类作为主要类目。所以，可以把《中图法》中的汽车理论、汽车设计和计算、汽车构造、汽车发动机、汽车材料、汽车制造工艺、汽车试验、汽车制造厂、各类型汽车、汽车驾驶与使用、汽车保养与修理、汽车燃料与润滑料等 12 个类作为 12 个分面，把所有汽车专业的词都对应到这 12 个分面中，将《中图法》原有的分类号给予每个词。这样，分类号－词对应表就基本编成了。这个表按分类顺序排列，供检索用。

词－分类号对应表的编制法是将分类号－词对应表的款目倒转过来，按词的字顺排列，供自动分类标引用。

在自动分类标引过程中，将从文献题名中自动抽出的词通过与词－分类号对应表核对，赋予《中图法》的分类号，建立分类号索引，提供分类检索途径。同一题名中的词因为分属于不同的分面，其分类号也就有多个。词仍应保留，建立关键词索引，提供主题检索途径。在对应表中如果能将等同关系词选定一种词形为正式词，其余为非正式词，提供非正式词转换成正式词的功能，则更好。

自动分类标引规则，是指从一篇文献获得的多个分类号中确定一个主要分类号的规则。确定主要分类号的依据，实际上就是图书馆中通行的文献分类规则（主要是所用体系分类法的类目内容范围划分规则）。根据通行的分类规则，将各个分面定出一个优先次序。一篇文献所获得的几个分类号中，哪个分类号所属分面在别的分类号所属分面之前，就确定那个分类号为主要分类号。如果遇有两个分类号所属分面都是最前的，则两个分类号均为主要分类号。对主要分类号必须加一个标志，以便在检索中必要时与其他次要分类号能区分开来。

用上述自动分类方法建立的检索系统，可以说是分类法主题法一体化的检索系统，其检索功能有：(1)可用单个分类号检索；(2)可用几个分面的分类号组配检索；(3)可对同一分面的分类号进行“逻辑和”检索或截号扩检；(4)可用单个词检索；(5)可用几个词组配检索；(6)可用词与分类号组配检索（有时需用截号检索法）；(7)可浏览选择分类号和浏览选择检索用词。

2.5 后控制词表

为自然语言检索系统配备后控制词表，是提高其检索效率的有效措施。

后控制词表的性质类似于入口词表，它是一种转换工具，是一种扩检工具，是一种罗列自然语言检索标识供选择的工具。

后控制词表的特点在于：其中的控制词（也可以是分类号）并非直接用于标引，而是对作为文献检索标识的自然语言词进行控制（建立等同、等级、相关关系）。因此，在后控制词表中，标引－检索用词是自然语言，非标引－检索用词却是人工语言，这与在一般词表中的情形正好相反。

后控制词表必须在检索系统中实有的自然语言检索标识的基础上进行编制（即必须以作为检索标识的自然语言原词为基础），以达到最大的覆盖率。否则将会大大降低其控制功能。这一点，是后控制词表的关键所在。

后控制词表较理想的结构模式是“分类词表＋字顺/轮排表”。

分类词表可设置下列等级：大类→小类→控制词→自然语言词。

后控制词表的编制过程包括两个阶段，即初始阶段和完善阶段。初始阶段是后控制词表编制成形，内容还不很丰富，但可以投入使用。完善阶段是在使用过程中词汇不断增补积累和体系不断调整细化。后控制词表的体系可用现成的分类表或词表作为框架，将关键词填入；也可对积累到一定数量的关键词进行归纳整理，形成系统。

后控制词表编制上的特点可归纳如下：(1)后控制词表是不断增长即不断增补积累而成的；(2)后控制词表的分类体系是随着词量的增加而逐步细化的；(3)后控制词表由于不用于文献标引，故其分类体系改变灵活，可进行较大的调整，对检索系统不会引起重行标引的问题；(4)后控制词表可以多种显示方式（也就是控制方式）并用。多种显示方式不一定要一步到位，可逐步增加；(5)各种后控制词表可根据实际情况选择显示概念关系的深入程度和显示的系统性。如有的只显示等同关系，有的显示等同关系和等级关系，有的显示等同、等级和相关关系；有的只有字顺显示，有的只有分类是

示，有的则用多种方式(包括词素轮排)显示；(6)后控制词表中的词间关系，应由人工判别确定。但可由计算机辅助，即利用字顺排列和词素轮排中的字面成族原理，用计算机寻找出可能的词族；(7)在后控制词表中，标引词与非标引词应有所区别(可用不同符号)。如果两者字面相同，应并存。

2.6 词素词表

汉语的词素词表(它不同于词根词表)首次见于"《军用主题词表》应用管理系统"。该系统具有自然语言入口功能。标引文献时，文献主题概念可全部用自然语言词自由表达。若表达文献主题概念的自然语言词与词表中的叙词一致，或与词表中的入口词(同义词和被组代词)一致，都可立即自动转换成叙词，自动将叙词登录人标引结果字段；若表达文献主题的自然语言词在词表中没有对应的叙词或入口词，该系统便会对自然语言词进行词素分析，利用词素相似性匹配原理，自动推荐一批含有相同词素的叙词供选择，通过人工判别，选定合适的叙词(包括若干叙词的组配)进行标引；若所推荐的词均不合适，则可将自然语言词作为自由词进行标引并同时作增补记录。

《军用主题词表》应用管理系统在应用词的相似度匹配原理时，以相同词素的个数为统计单位，并结合叙词词素的位置特征(如词素在词尾、在词首、在词中)及长度特征进行加权，可调整权值来扩充或压缩推荐词的数量以方便选择，并加人同义词素避免遗漏等，从而使所推荐的词更有针对性和全面性。这种方法，在无形中提高了词表的入口率，无疑使标引工作更为容易。该系统所用的词素调表(称为知识库)采用在叙词表自身词汇和语义关系的基础上，进行自动、滚动切分，辅以少量人工干预的方法，切分效率较高。

2.7 网上资源检索工具

网上资源检索工具通常称为搜索引擎，是在环球网产生后伴随着人们在网上快速检索知识－信息资源的需要而产生的新生事物。据报道，目前在因特网上的较大型的中文搜索引擎已有40种以上，外文的搜索引擎则数以百计。这些网上资源检索工具一般提供分类检索途径、专题检索途径和关键词检索途径。它们实际上是一种网页网址检索系统，其数据库中收录有几十万乃至几千万个网页网址。这些检索工具虽已在网上资源的检索中发挥了很大作用，但仔细考察，它们的编制还是很粗糙的，并不易用，还有待改进和提高。特别是专业性的检索工具还需大量编制。网上知识－信息资源大多未用情报检索语言标引，所以，网上资源检索工具必然是一种对自然语言检索进行控制的工具。目前对这类检索工具作深入研究的文献还不多见，是有待在编制过程中大力加强研究的。

2.8 各种自然语言词表的相似性和可灵活使用

上述自然语言接口用对应表、自动抽词词典、自动赋词用对应表、自动赋号用对应表、自动分类用对应表、后控制词表、词素词表等，其功能有相似性，因而可以灵活使用，可以互相取代，可以"组装"和"拆卸"。例如，自然语言接口用对应表主要用于检索但也可用于半自动赋词赋号标引(机助标引)，一部分自动赋词赋号标引用对应表具有自然语言接口功能，自动抽词词典与自然语言接口用对应表联接可取代自动赋词赋号用对应表，把自动赋词赋号用对应表款目倒过来，可改造成后控制词表的框架，具有自动赋词赋号双重功能的对应表在实际标引中可只取其一种功能，但在检索中却可两种功能都用，后控制词表也可与自动抽词词典联接，等等。

任何正规的分类表和词表检索系统，都可以增加自然语言接口，而不会影响原有的标引工具和标引数据，只会增益，有利无弊。

任何自然语言检索系统一旦使用某种自然语言词表，都是得到某种程度的控制，可提高检索效率，提高易用性。

由此可见，大量编制自然语言词表可使文献标引－检索用语言推向一个新的发展阶段，是大有可为的。

2.9 为《中国分类主题词表》配备自然语言词表问题

《中国分类主题词表》包含了我国两大主要情报检索语言语种而且是互相对应的，因此可以说已成为最主要的标引工具。它向自然语言化发展，无疑具有重大意义。

《中国分类主题词表》该采取哪种自然语言化的方式呢？或者说，为它配备哪种自然语言词表呢？

我以为，可为《中国分类主题词表》配备自然语言接口用对应表。自然语言接口用对应表既可用于检索，也可用于半自动赋词赋号标引（机助标引）；增补非关键词以后，则可用于自动抽词；将其对应款目倒过来，也可作为某些专业的后控制词表的框架和基础。

《中国分类主题词表》的某些类目可配备自动分类用对应表，用于论文的自动分类。

《中国分类主题词表》配备自然语言接口用对应表和自动分类用对应表以按专业分别编制为好。某个类编到一定规模，即可投入使用。要为整个分类主题词表配备完整的对应表是一项旷日持久的工程，损失时间太多，是不可取的。况且，这类对应表是要不断增补、逐步积累的，决不可能一次编制完善。对应表只需机读版，供联机使用，增补修订也比较灵活。

3 自然语言词表编制中的方法和技术问题

3.1 编表策略

上面曾提到，专业性的抽词词典较易编得好，而且专业范围越小越易编得好。其实，各种自然语言词表都是如此，应当分专业编制。综合性的自然语言词表过于庞大，不但难于编制，而且使用起来效果也不会很好。

自然语言词表都是不断积累、逐步完善的。编到一定规模（不必很完整），即应投入使用，在使用中增补、完善。

自然语言词表都是联机使用的，只需要机读版，没有必要出印刷版。

编制工作分散比集中好，应发挥各个需要的单位的积极性。

各单位所编的各种自然语言词表，最好有一、二个单位负责集中，进行交流，以充分发挥其社会效益，促进我国图书情报事业的发展。

3.2 词汇收集

自然语言词表的词汇是否丰富，是决定其功能强弱的主要因素，当然是越丰富越好。收集词汇不要怕重复，重复的词可用删重命令删除。

自然语言词表的词汇应包括：(1)正规的术语。这部分语词可从各种术语词典、检索词表和分类表（特别是机读词表和分类表）中收集；(2)非正规的语词。这部分语词见于文献，在上下文中或在专业范围内可理解其涵义，但当脱离上下文或专业范围时，就涵义不清了。这类语词往往不见于词典，语言学界也不认为是正规的词，或认为是不符合语法的词，但这些词也是有用的。这些词在开

始阶段可通过切分一定数量的文献题名来收集，以后再在使用中不断积累。

如果要对文本进行“无剩余抽词”，则抽词词典应包括关键词和非关键词。非关键词是各专业的抽词词典通用的。

3.3 词汇分面化组织

将收集到的词汇进行分面化组织，对各种自然语言词表的编制都很有帮助。经过分面化组织，就可比较容易地将自然语言语词与情报检索语言语词进行对应。

利用字面成族原理，使用词素匹配检索方法，是集中同族词的简便措施。同族词一般属于同一分面。

3.4 对应规则

除自动抽词词典外，各种自然语言词表的编制工作中最重要的内容就是标识或语词以概念为基础的对应转换。对应转换是否准确，是自然语言词表质量的决定因素。

编制对应表时应注意：

⑴列入对应表的语词，其对应关系应是确定无疑的。若不能肯定其对应关系的准确性，最好不要列入对应表；

⑵对应的范围可随需要而定。如只包括同义关系，还是扩大到近义关系、专指与泛指的关系，等等；

⑶对应表源词字段可不包括词表正式词，但最好包括词表正式词(即词表正式词也作为一个自然语言词重复列入对应表，加入“词表正式词→词表正式词”款目)。这种对应表的编制法，是将原有词表的款目词复制出来，填充到源词字段(被转换的词)和目标词字段(转换成的词)，并与其他对应条目混合排序即可。

3.5 计算机在编表中的应用

在各种自然语言词表的编制中，下列作业可利用计算机：

⑴利用文字处理软件用人工切分文献题名，然后套入数据库，用删重命令进行整理的方法来收集词汇。有关学科、专业的词表如有机读版，其中的词可用套录的方法来收集；

⑵对收集来的词汇，用词素匹配检索的方法集中同族词，进行分面化组织；

⑶对词汇进行自动排序或对词表进行改组；

⑷编辑修改工作尽量在计算机上进行，最后完成的词表是一个词表数据库(机读版)。

参考文献：

[1]张琪玉.情报语言学基础(增订二版).武汉大学出版社，1997

[2]张琪玉.情报检索语言的发展趋势(与吴建中的对话).图书馆杂志，1996(4)

[3]张琪玉.世纪之交中国情报语言学发展之路.1997理论学术年刊(图书馆杂志1997年增刊)

[4]张琪玉.情报语言学领域亟待研究且潜藏较富的研究课题.发表中

[5]张琪玉.自然语言在情报检索中的应用.情报理论与实践，1996(3)

[6]张琪玉.自然语言检索中各种因素对检索效率的影响.情报理论与实践，1997(5)

[7]张琪玉.自然语言与人工语言的对应转换——情报检索语言走向自动化之路.中国图书馆学报，1996(1)

[8]张琪玉.情报检索语言走向自动化之路与《中图法》发展新目标.北京图书馆馆刊，1996(4)

[9]张琪玉.概念或标识自动转换技术的应用.图书馆杂志，1998(6)

[10]张琪玉.分类法主题法一体化自动标引系统的基本原理和方法.图书馆论坛,1995(6)
[11]张琪玉.文献题名自动抽词－分类标引系统.图书馆杂志,1998(4)
[12]张琪玉.人工语言与自然语言、先控制与后控制的界限在计算机系统中可淡化或取消.图书馆杂志,1997(5)
[13]张琪玉.一个优秀的叙词机助标引软件.图书馆杂志,1999(1)
[14]张琪玉.缺乏抽词词典是自动抽词标引难以普及的主要原因.图书与情报,1998(2)

论网络环境下自然语言在情报检索中的地位与作用

湖南省科技信息研究所　黎盛荣

[摘要]本文首先在回顾情报检索语言发展过程的基础上,对网络环境下自然语言在情报检索中的地位与作用进行了系统地分析和论述。

[关键词]情报检索语言　分类法语言　主题法语言　自然语言　计算机网络

随着现代高新信息技术的不断发展,情报检索出现了联机检索和光盘检索系统,使传统的手工检索方法进入了高速准确和自动化的新阶段。特别是全文数据库的问世,有力地促进了情报检索语言的发展与革新,分类法语言和主题法语言开始向机编化、机读化、一体化和网络化方向迈进。同时,曾一度被冷落了的自然语言又开始受到重视,并成为适应计算机情报检索网络化发展要求的情报语言,成为人工智能技术和自动标引不可缺少的基本要素之一。因此,自然语言(关键词)在计算机情报检索中具有重要的地位与作用。

1　情报检索语言发展概况

情报检索语言发展史,源远流长,如果从我国分类法语言看,可以追溯到西汉时期,即公元前1世纪末,刘歆编辑的《七略》是我国古代分类法语言的典范。

公元前250年左右,亚历山大图书馆卡利马考斯曾编辑了一部类似图书分类法的书,这是国外分类法语言的萌芽。

1545年,瑞士吉士纳编辑了一部《万象图书分类法》,这是欧洲第一部图书分类法,其内容分字学、数学、修养和高等科学四大部类,21个大类,它为国际图书分类法的发展奠定了良好的基础。

1818年,美国学者普尔(W.F.POOL)首次提出了《杂志和连续出版物的字顺主题索引方法》,它标志着西方主题法语言的开端。

1952年,罗杰特(M.Roget)编辑的一部《Thesanras of English Words and Phrases》标题表,是国外最早的主题词表。

1954年,美国波士顿梅堪特图书馆编印的字典式目录和1876年美国克特(C.A.Cutter)发表的《字典式目录条例》(Rule for a dictionary Catalogue)把主题标识、著者标识、书名标识三者的字顺结合起来,这实质上是主题法语言、著者语言和书名语言三结合的检索标识系统,是人工语言与自然语言兼容化的初步尝试。

19世纪初,国外图书分类法语言开始初步形成自己的基本体系。1876年,《美国杜威十进分类法》(简称"DC"或"DDC")第一版问世了。1905年,比利时目录学家奥特勒(P.Otlet)和拉封登(H.Lafontaine)编辑了一部《国际通用文献分类表》。类目有33000条,字顺索引38000条。1927年改名《国际十进分类法》,简称"UDC"

20世纪初,《杜威十进分类法》传入我国,当时,仿杜威十进分类法的沈氏分类法、杜氏分类法、

刘氏分类法、皮氏分类法等先后出版发行。这些图书分类法在我国图书分类学史上产生了深远的影响。

1933年，印度著名图书馆学家阮冈纳赞（S.R.Ronganathan）发表了《冒号分类法》，简称“CC”。其特点是提出了一种新型的主题分析方法，即分面组配法。它在分类法理论上形成了独特的组配式分面体系，这在分类学和情报检索语言学领域是一个创举，具有开拓性的历史意义。

1934年，我国图书馆学家沈祖荣先生的《标题总录》问世。1937年，吕绍虞先生的《标题表》也面世。随后，程长源先生的《中文标题表》出版了。

20世纪中叶，随着微电子学的问世和计算机技术的发展，产生了主题法语言。

1951—1952年，陶布（M.Taube）和古尔（C.D.Gull）等人提出了单元词组配索引法.不久，美国穆尔斯（G.N.Mooers）发明了叙词法，它是在传统的标题法、元词法、键词法的基础上发展起来的一种新型的情报检索语言。

自然语言检索，在我国始于50年代末和60年代初。但情报检索语言在我国进行深入地研究和应用，是在80年代才形成高峰。

从50年代开始，国内先后出版了《中国人民大学图书馆图书分类法》、《中国科学院图书馆图书分类法》、《中国图书馆图书分类法》等30部分类法，从而使我国图书分类法语言获得了巨大发展。

60年代，我国主题法语言的研究才起步。70年代中期，“748工程”开始启动，作为该工作配套项目的《汉语主题词表》于1975年开始编辑，1980年正式出版。

80年代，由于计算机情报检索在图书情报工作中的广泛应用，我国对分类法语言与主题法语言一体化的研究从理论探讨进入了实际编表阶段。《中国分类主题词表》于1986年开始编辑，1994年正式出版发行，它是一种新型的分类主题兼容性的情报检索语言，标志着中国情报检索语言研究迈向了一个新台阶。

90年代，张琪玉教授的《情报语言学基础》一书出版了，它作为一种新的学科学派在我国图书馆学和情报学界产生了较大的影响，对推动中国情报检索语言的发展具有一定的现实意义。

广东省中山图书馆在《中国分类主题词表》机读数据的基础上，成功地研制了计算机辅助文献分类标引、主题标引与检索系统，并在本地区公共图书馆系统应用，效果很好。

40多年来，我国编制的图书分类表有130多部、主题词表120多部、发表的论文6000多篇、著作100多种。从事研究的人员达2000多人，其中最有影响的代表人物有沈祖荣、杜定友、刘国钧、皮高品、张琪玉、白国应、刘湘生、侯汉清和戴维民等四代学人。

总之，20世纪以来，中国情报检索语言在体系分类法的基础上有了巨大的发展，开拓了主题法、分类主题一体化、引文索引法、新的代码法、序号法、自动标引和全文检索系统等各个领域。

2　自然语言在情报检索中的地位与作用

自然语言在我国情报检索工作中的应用，始于50年代末和60年代初。从分类法语言、标题法语言、单元词语言到叙词语言的演变过程，也反映了自然语言不断发展的基本规律。特别是随着计算机网络化的发展，使自然语言成为情报检索语言研究的重要课题之一。

所谓情报检索语言，张琪玉教授在《情报语言学基础》一书中指出：“情报检索语言是根据情报检索的需要而创制的人工语言”。它又称检索语言、标引语言和情报语言等。而自然语言是反映文献题名、文摘和全文内容特征的原词，它是通过文本中的概念来揭示文献内容的关键词语言。自然语言的检索方法就是对文献本身的用词进行直接处理的全过程。由此可见，自然语言是一种没有

通过规范化的情报检索语言。其特点是：

(1)自然语言是作者撰写文本的原词，直观性与专指性强；

(2)自然语言是反映文献题名、文摘和全文内容的原用词，检索途径多；

(3)自然语言便于计算机自动抽词，不用人工标引，省去了查分类主题词表的时间；

(4)自然语言能适应科学技术发展的动向，随时可以增补新词；

(5)自然语言检索方便，查准率高；

(6)自然语言标引速度快，对标引员的要求不高，容易掌握。

最近，张琪玉教授提出了一种新型的情报检索语言，即“学科－事物概念组配型检索语言”模式。这种语言实质上是一种分类主题兼容性检索语言。而检索语言的兼容性是情报检索在网络环境下提高检索效率和实现资源共享的重要条件。

由于计算机容量大，运行速度快、检索功能强等特点，所以使自然语言重新受到重视，大有用武之地。如美国 DIALOG 系统和 ESA－Quest 系统的数据库中，自然语言就占有相当高的比例。

美国俄亥俄州大学图书馆馆长、美籍华人李华伟博士曾预言：“未来的情报检索语言是以自然语言为主的发展方向”。

但是，自然语言并不是十全十美的情报语言，它也存在一定的局限性。

其一，自然语言近义词、同义词、多义词及其相关的词未能进行规范化。再者，由于汉语语音缺乏科学性，所以影响到查全率。

其二，自然语言如何准确地揭示文献的具体内容和检索课题有效地匹配问题，是值得研究的重要课题。

由此可见，电子计算机是很难区别哪些是有用的词、哪些是无用的词。特别是汉语本身语义关系的复杂性，是计算机难以识别的。因此，利用自然语言检索文献信息和自动标引，并达到资源共享的目的，首先要解决好汉语自动切分的难题，这在情报检索语言研究领域称为汉语分词技术。

我国在汉语分词技术研究方面，有不少专家和学者进行了大量的试验与探索。其中具有代表性的如陈培久的“词典切分组词标引法”、王永成的“词部件组配标引法”、吴蔚天的“字分类组词标引法”、杨学山的“规则词典标引法”、张琪玉的“文献题名自动抽词——分类标引系统”等。

文献题名自动抽词—分类标引系统是由三个子系统组成，具体内容如下：

①自动抽词子系统；

②抽词词典增补子系统；

③检索子系统。

自动抽词子系统包括以下五个文档：

①题名库；

②词典库，是系统的核心库；

③词块库，是过渡文档，便于与题名匹配；

④索引库；

⑤待处理库。

在文本中抽取关键词进行标引和检索，虽然查全率不高，但其查准率相对面言还是较高的。因此，从文献题名中取提关键词作为情报检索标识，在自然语言检索系统中具有重要的意义。

本世纪 90 年代，美国把自然语言处理技术应用到情报检索领域中，真正实现了非布尔逻辑的自然语言检索系统。目前，具有一定代表性的检索系统有美国西部出版公司开发的 WIN 系统、FREESTYLE 系统、MNIS 系统等。

随着计算机技术、通信网络技术和多媒体技术的发展，促进了情报检索语言的不断更新，具体类型概括起来有以下几种：

(1)人工语言(受控语言)

①分类法语言(体系分类法、分面组配法)；

②主题法语言(标题法、元词法、叙词法)；

③代码语言(语义代码、化学代码)。

(2)自然语言(非控语言)

①关键词语言(含自由词)；

②题名语言(书名、刊名、篇名、题录)；

③著者语言(含译者)；

④序号语言(标准号、专利号、报告号)；

⑤引文索引语言；

⑥全文语言。

(3)兼容语言

①分类法与主题法兼容语言；

②关键词与自由词兼容语言；

③叙词与自然语言兼容语言。

张琪玉教授在“世纪之交中国情报语言学发展之路”一文中提出：“理想的情报语言，应是学科聚类系统与事物聚类系统的结合、先组式语言与后组式语言的结合、体系分类法与组配分类法的结合、人工语言与自然语言的结合、号码标识与语词标识的结合、系统序列与字顺序列的结合、不变概念代码与可变概念体系的结合”。张琪玉教授的这一设想为我国情报检索语言研究指明了方向。但从理论上看，其实质和杜定友、刘国钧两位学者早在60年代提出的“走向新型的分类主题目录的发展道路”和“分类法与主题法相互取长补短、相互结合”的观点没有多大的区别，只不过在前人的基础上作了更进一步的综合与深化。

笔者认为：研究汉语语法分析和语义分析的高智能情报语言系统，是我国21世纪情报检索语言研究的重要课题，关键是在于首先解决好汉语切分技术问题。为此，建议有关部门组织计算机专家、情报学专家和语言文字学专家共同协作攻关。这样才能使中国情报检索语言取得突破性的进展，迈向一个新台阶，为开创我国情报检索工作现代化、网络化的新局面作出贡献。

参考文献：

[1]张琪玉.文献题名自动抽词分类标引系统.图书馆杂志，1998(4)

[2]白国应.中国检索语言发展方向.理论学术年刊，1997年，上海图书馆杂志社

[3]赵英莉.新信息环境下的自然语言检索与我国对策研究.图书馆杂志，1998(3)

[4]宋明亮.情报语言学研究的新领域.图书情报工作，1994(5)

[5]周全明.我国情报语言学研究的方向重点与热点.图书与情报，1994(1)

[6]洪漪.我国信息网络建设中的检索语言.中国图书馆学报，1995(3)

[7]傅日.试论词表兼容性.情报学报，1990(6)

[8]黄邦高.论检索语言的新概念.图书情报知识，1998(3)

[9]黎盛荣.我国自动标引研究的现状及其发展趋势.图书馆论坛，1995(1)

《全国报刊索引》标引工作应用自然语言的设想

王越倩

[摘要]本文提出了《全国报刊索引》标引工作应用自然语言的设想，分析了人工抽词以及建立自动标引、自动分类系统的理论基础和可行性。具体阐述了《关键词—主题词对应表》和《主题词—分类号对应表》的建立方法以及对自然语言标引进行后控制的方法。

[关键词]《全国报刊索引》 关键词 自然语言标引 自然语言检索 自动标引 自动分类 后控制词表

自然语言化是当代检索语言发展最重要的特征和趋势。目前，自然语言在情报检索中应用的研究很活跃，尤其在自然语言接口与人工语言结合使用方面已有很大成就。本文就以《全国报刊索引》的标引工作为例，提出自然语言应用的设想。

1 《全国报刊索引》的发展概况

《全国报刊索引》创办于 1954 年。它收录国内公开和内部发行的全国性、专业性、省、市、自治区及部分地方的中文报纸和中文期刊，撷取重要资料，由上海图书馆编辑出版，分为哲学社会科学版和自然科学版。它是目前全国收录社科期刊数量最大，历史最悠久的一本题录性检索工具，其学科覆盖面广，收录期刊和报纸全面，著录准确，标引详细，提供分类、著者以及刊名等检索途径，受到广大用户欢迎。

《全国报刊索引》自问世以来，工作流程日趋合理、精简，主要体现在标引工作效率的提高上。社科版由原来单纯的分类标引到分类标引与题中人名标引并用，发展到目前的分类标引与主题标引相结合的方法。用《中国图书资料分类法》(第三版)进行标引、排版，结合《汉语主题词表》进行主题标引。由于使用了对应标引，工作速度大大提高，由原先的每天数十条提高到 100 条/每天。科技版则使用关键词进行主题标引，直接使用作者标引的关键词加以修改增补，从很大程度上提高了工作效率，尽管如此，由于文献量的不断增加，科技版以及《中文社科报刊篇名数据库》的上马，现有的方法已力不从心。于是我们力求寻找一种更为有效的检索语言和标引方式。

2 《全国报刊索引》应用自然语言的设想及可行性分析

美国图书馆学家兰开斯特提出："由于智力加工所需成本的不断上涨，计算机存贮费用的不断下跌，以机读形式存取的文本(包括电子邮件，电子报刊)数量的逐渐增多以及在联机检索中对熟悉的'中间人'依赖的逐渐减少，自然语言将变成情报检索的规范，普通受控叙词表的使用将会衰退"。[1]随着计算机技术的飞速发展，推动了情报检索的计算机化，使文献标引和检索趋向自动化，使自然语言标引及检索得以实现。情报检索所用的语言已不再局限于人工语言。可见自然语言的

应用和自动化将是情报检索语言的发展趋势，也将是《全国报刊索引》的发展方向。

根据《全国报刊索引》编辑工作的实际需求以及情报检索语言的发展趋势，我提出这样一个设想：能否将自然语言与自动化结合起来，应用自然语言，自动抽取关键词，由计算机转换成为主题词和分类号呢？从而实现计算机辅助分类标引和主题标引。

《全国报刊索引》应用自然语言的核心问题是自动抽词和自动标引系统的建立，我们选用《新疆畜牧文献数据库》作为专业文献的来源，选择其中的饲料(S816)类作为试验对象，就这一设想的理论基础以及实际应用进行可行性分析。

2.1 自动抽词的理论基础与应用

自然语言检索用词一般取自文献本身(题名、摘要、各级小标题、全文)、个别由标引员自主赋予(即自由标引)。相对于情报检索语言来说，使用自然语言的优点是：

(1)不必查看词表，降低标引工作的难度和成本，提高标引工作的速度；

(2)直接使用文献用语，可以改善标引的专指性和一致性，从而提高检索的效率；

(3)用户熟悉自然语言，使用起来方便得多，尤其在联网环境中；

(4)词汇更新快。

考虑到《新疆畜牧文献数据库》中(S816)属科技文献，其题名具有很强的报道性。根据S816类随机抽样的记录统计，有49%的关键词直接从题名抽取，20%的关键词与题名中的词有等级关系、相关关系，33%的词须由人工修改或添加。由于关键词能够较好的表达文献内容，因而可以选择关键词抽词标引，对一篇文献标引3至5个关键词。

目前，关键词自动抽词标引通常使用大辞典匹配的方法：用计算机对题名进行自动切分，通过《关键词词典》及《禁用词词典》来匹配。其自动化程度高，标引速度较之人工标引可提高十几倍，标引的吻合率极高。关键词自动抽词的质量取决于《关键词词典》和《禁用词词典》的准确性和完备性。因此，《关键词词典》的编制是一个关键技术。

2.2 自动标引的理论基础与应用

自动标引是自然语言的核心问题。《全国报刊索引》采用主题和分类两种语言标引，从这个意义上讲，自动标引系统的实现包括两个层面：(1)主题词的自动标引，自然语言与受控语言的转换，即关键词与主题词的对应与转换(2)分类号的自动标引，主题语言与分类语言之间的对应转换。

主题词自动标引的实现依赖于《关键词—主题词对应表》，由人工将每个关键词转换成主题词。由计算机根据对应表将文献中的关键词转换成主题词。并由人工对转换的结果进行判别。

由此可见，无论是自动抽取关键词还是自动标引系统的建立都有可靠的理论作为后盾，而且在应用上是可行的，但实现的程度要取决于几部词表的完善程度以及计算机水平的提高。

3 两部词表的编制方法

3.1 《关键词—主题词对应表》的编制

《关键词—主题词对应表》的编制以《中国分类主题词表》中的主题词为规范进行转换，并参照《汉语主题词表》，转换时遵循以下步骤：

(1)将关键词进行适当的处理，含义相关的词合并，含义重叠的词切分，每一篇文献生成若干个

关键词或关键词串。

例1:籽粒苋　青绿饲料合并为:青绿饲料—籽粒苋

例2:鱼粉贮藏　饲料贮存合并为:鱼粉—饲料贮存

例3:尿素　石灰　麦秸处理　处理为:尿素—麦秸—秸杆处理

石灰—麦秸—秸杆处理

(2)逐次将关键词(串)转化为主题词(串),在转化的过程中要注意识别同义词、准同义词、近义词,将其归入相应的主题词下,对应形式大致有以下几种:

1.一对一。如:塑料袋青贮—>青贮

2.一对二。如:颗粒饲料标准—>颗粒饲料—标准

3.一对多。如:配合饲料厂—>饲料加工—工厂

4.二对二。如:木薯—消化率—>粗饲料—消化率

5.二对多。如:鱼粉鉴别—镜检鱼粉—>饲料—检验—方法

(3)制作词表,将关键词依字顺排列,词表的格式如表1:

表1 《关键词—主题词对应表》(片断)

关键词	主题词
氨化秸杆	藁杆—调制
大豆饼	豆类作物—精饲料
大豆粕	豆类作物—精饲料
大豆饲料	豆类作物—精饲料
多汁饲料	多汁饲料
秸杆氨化	藁杆—调制
秸杆处理	藁杆—调制
麦饭石—天然饲用矿产	矿物质饲料—麦饭石
青贮高粱	高粱—青贮
青贮玉米	玉米—青贮
热喷秸杆	藁杆—调制
饲料贮存	饲料贮藏
饲料质量—定量分析	饲料—分析
苏联标准—饲料粒度标准	饲料标准—苏联
塑料袋青贮	青贮
塑料袋青贮饲料	饲料—青贮
特殊饲料	特种饲料
鱼粉—饲料贮存	鱼粉—饲料—贮藏

在《关键词—主题词对应表》的制作过程中要注意以下几点:

(1)由于词表是按饲料(S816类)单学科建立的,具有专业词表的性质,因此,非饲料类学科的关键词不收入其中。

(2)转换中以《中国分类主题词表》中的主题词为规范,根据实际需要可以增加对应的主题词及词串。如S816类下,没有涉及饲料资源和饲料原料的主题词,导致一些关键词无确切对应的主题词。因此,在S816类下人为添加"饲料资源"和"饲料原料"两个主题词。一些使用频率很高的专有名词也可以添加作为主题词。

(3)注意同义词、准同义词、近义词的归并。例如表1中:"秸杆氨化"、"氨化秸杆","秸杆处

理","秸杆利用","热喷秸杆"都归入"藁杆—调制"下。"大豆饼","大豆粕","大豆饲料"等都归入"豆类作物—精饲料"下。

3.2 《主题词—分类号对应表》的编制(表2)

主题词—分类号的转换是一个双向的过程:建立《主题词—分类号对应表》的同时建立《分类号—主题词对应表》,以便于读者从主题和分类两个角度进行检索。

使用《中国分类主题词表》作为分类号和主题词之间兼容转换的工具,通过标引数据同时完成文献的分类标引和主题标引。

类号标引的原则与方法决定了对应表的质量。因此在标引时应力求标引深度一致,必要时可以加细分类表,并遵循专指标引—上位类标引—靠类标引的优先顺序。由于一篇文献进行分类标引后,往往会有一个以上的分类号出现,而《全国报刊索引》的标引要求是使用一个分类号,这就要建立一个自动分类规则,即从一篇文献获得的多个分类号中确定一个主要的分类号。根据通行的分类规则及实际需要建立以下的规则:

(1)若几个类号相同则合并;

(2)若类号是上下位类关系则取其上位类;

(3)若论述事物及事物的某一方面,则归入事物类;

(4)若论述的重点有偏重,则归入重点类目;

(5)论述的事物无偏重,则随机抽取一个类目。

表2 《主题词—分类号对应表》(片断)

主题词	分类号
动物性饲料—鱼粉	S816.48
藁杆—调制	S816.53
青贮饲料	S816.5
饲料资源	S816
饲料贮藏	S816.9
玉米—青贮	S816.9
猪血发酵饲料	S816.6

《主题词—分类号对应表》的编制按主题词的字顺排列,《分类号—主题词对应表》将其逆转过来,并按类号归类排列,一号可以对应双词或两个以上的词。

表3 《分类号—主题词对应表》(片断)

分类号	主题词
S816	多汁饲料;蛋白质饲料;特殊饲料;膨化饲料浓缩饲料;混合饲料;特种饲料;人工饲料
S816.1	饲料—标准;饲料粒度标准;饲料—标准—苏联
S816.42	豆类作物—精饲料;大豆饼;大豆粕;大豆饲料;大豆粉
S816.53	藁杆—调制,稻—藁杆—调制;麦—藁杆—调制;藁杆—氨化;藁杆—热喷
S816.71	矿物质饲料;矿物质饲料—食盐;矿物质饲料—钙;矿物质饲料—磷;天然饲用矿产—沸石
S816.79	饲料添加剂—消化药;消化酶;植酸磷分解酶饲料酵母
S816.9	饲料—处理;饲料—贮存;青贮;高粱—青贮;玉米—青贮;塑料袋青贮;棉粕分级

4 后控制词表的编制

《全国报刊索引》应用自然语言具有许多优点：标引容易，标引速度快，检准率也有一定的保证等。但是“任何以语词作文献检索标识的自然语言检索法，检全率都存在着检索者构造检索策略困难和检全率较低的问题”，而且“自然语言检索法表达概念较为自由，存在着对大量的同义词、近义词、同义词组、近义词组缺乏控制的弊病”。除此之外，自然语言还存在着对词量、词形、词间关系不进行控制的缺点，从而影响检索效率。但若采用后控制词表对自然语言系统进行事后控制，那么以上的问题可以在很大程度上得到解决，大大提高检索效率。

后控制词表的编制有多种方式，考虑到先前已经编制了《关键词—主题词对应表》并在一定程度上集中了同义词、准同义词，所以考虑将原先的关键词进行轮排，(见表4)编成一个题内(KWIC)型轮排索引，用户可从入口词的某一个词素出发，按照字顺找到有关的入口词及对应的主题词和分类记号。整理为入口词表，不失为后控制词表编制的有效方法。

这种题内关键词索引采用《汉语主题词表》(自然科学增订本)轮排索引的编制技术，其中包括顺序与逆序排列结合的双向排序法；即先从排检位置出发，由左向右顺序排列词素，然后再由右至左逐字逆位于排检位置左侧的词素。这种排列符合汉语语言特点，可以把词根相同的入口词集中，显示族性相关的入口词，而且可以把有关的某一事物各个方面的入口词集中。因为这种排序法可以进行检索词等级关系显示和方面关系(即相关关系)显示。

上述后控制词表编制方法简便，它显示了词间关系，对同义词、同义词组及准同义词进行了控制，克服了应用自然语言的弊病，大大提高了检索效率。

表4 后控制词表的款目形式

	——排检位置 ↓			
大豆	饼	Y	豆类作物—精饲料	S816.42
大豆	粕	Y	豆类作物—精饲料	S816.42
秸杆	处理	Y	饲料—贮藏	S816.9
	大豆饼	Y	豆类作物—精饲料	S816.42
	大豆粕	Y	豆类作物—精饲料	S816.42
	大豆饲料	Y	豆类作物—精饲料	S816.42
	豆类作物—精饲料	D	大豆饼；大豆粕； 大豆饲料	S816.42
	多汁饲料	Y	多汁饲料	S816
	多汁饲料	D	多汁饲料	S816.42
青贮	高粱	Y	高粱—青贮	S816.9
	藁杆—调制	D	秸杆处理；秸杆利用； 热喷秸杆	S816.53
热喷	秸杆	Y	藁杆—调制	S816.53
	秸杆处理	Y	藁杆—调制	S816.53
	秸杆利用	Y	藁杆—调制	S816.53
	高粱—青贮	D	青贮高粱	S816.9
秸杆	利用	Y	藁杆—调制	S816.53
	青贮高粱	Y	高粱—青贮	S816.9

	青贮饲料	Y	饲料—青贮	S816.9
	青贮玉米	Y	玉米—青贮	S816.9
	热喷秸杆	Y	藁杆—调制	S816.5
饲料袋青贮	饲料	Y	饲料—青贮	S816.9
大豆	饲料	Y	豆类作物—精饲料	S816.42
多汁	饲料	Y	多汁饲料	S816
	饲料贮存	Y	饲料—贮藏	S816.9
	饲料—青贮	D	青贮饲料;塑料袋青贮饲料	S816.9
	饲料—贮藏	D	秸杆贮存;饲料贮存	S816.9
	塑料袋青贮饲料	Y	饲料—青贮	S816.9
	玉米—青贮	D	青贮玉米	S816.9
秸杆	贮存	Y	饲料—贮藏	S816.9

5 小结与问题

根据《全国报刊索引》应用自然语言进行计算机辅助标引和分类,我们可以总结出大致的工作流程(见图1):

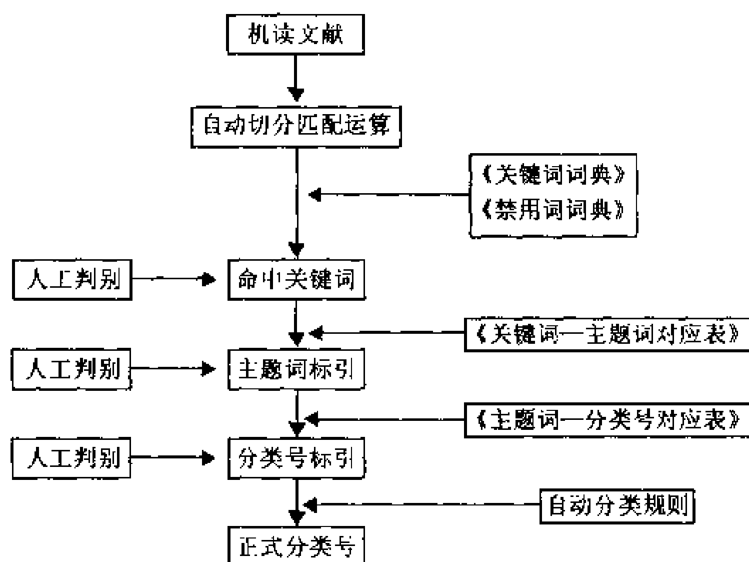


图1 工程流程

例如,现有一篇题名为"浅谈我国秸杆饲料资源的开发利用"的文献,现将用计算机标引过程总结如下:

1.将机读形式的题名用《关键词词典》与《禁用词词典》匹配,获得其中关键词:"秸杆饲料"、"饲料资源"、"开发利用",由人工加以判别,增加关键词"粗饲料"。

2.对关键词进行处理,加工成:"秸杆饲料—粗饲料"、"饲料资源—开发利用"。由计算机按照《关键词—主题词对应表》将关键词串转换为正式主题词;"粗饲料"、"饲料资源—开发利用",并辅以人工判别、增删或修改不准确的主题词。

3.由计算机将主题词按《主题词—分类号对应表》转换为 S816.5、S816 两个类号,并辅以人工

判别，增删或修改不准确的分类号，根据自动分类规则，确定最终的类号为S816。

由此可见，《全国报刊索引》标引工作应用自然语言的两个特点：

1. 自然语言应用要走自然语言＋后控制的道路。

2. 自然语言的应用要走自动化的道路。

所以，怎样减少人工辅助、提高自动化的程度及后控制词表的编制将是自然语言应用要解决的问题和努力的目标，在实际操作中，交叉学科的处理，自动抽词的实际水平，词表的规范和维护，这些问题都有待于解决，但是我相信，自然语言的应用将是《全国报刊索引》发展的目标。

参考文献：

[1]（美）F.W.兰开斯特著；侯汉清，戴维民译.情报检索语汇控制.同济大学出版社，1992
[2]侯汉清.索引技术和索引标准.北京图书馆出版社，1997
[3]赖茂生，王延飞.计算机情报检索.北京大学出版社，1993
[4]《中国图书馆图书分类法》编委会.中国分类主题词表（第一卷）.北京：华艺出版社，1994

档案信息高速公路——情报检索语言在档案界发展的又一机遇

空军政治学院信息管理系　张正强

[摘要]本文从各国档案界情报检索语言的运用分析入手,仔细分析了美国、加拿大和澳大利亚的档案在建立档案信息高速公路时对情报检索语言的关注,并深入地分析了其中的逻辑成因,指出信息的标准化是情报检索语言在档案界发展的重要前提、信息的规模化利用是情报检索语言在档案界发展的本质要求、图书馆界在创制和使用情报检索语言方面的成功经验是情报检索语言在档案界发展的重要借鉴。并结合我国的实际提出了我国档案界进一步发展运用情报检索语言的策略。

[关键词]情报检索语言　档案信息高速公路　美国　加拿大　澳大利亚

随着世界范围内的档案信息高速公路建设的兴起,使得情报检索语言在档案界的应用又出现了高潮。因为在建设档案信息高速公路时,面向信息主题和信息内容的检索成为档案检索中的关键,由此,档案界的目光又一次集中到情报检索语言。而且在世纪之交,这高潮来得如此之猛烈,范围如此之广泛,充分展示了情报检索语言在档案界的生命力及在未来的信息社会中的地位与作用。

1　各国档案界情报检索语言的运用

1.1　美国档案界情报检索语言的运用

美国档案界广泛地运用情报界语言是在本世纪的 80 年代末和 90 年代初,随着美国档案信息系统集成到国家信息系统、档案信息高速公路的崛起而发展起来的。在此之前,美国档案界也存在着对情报检索语言在档案中的运用有不同看法,其主要的观点是,档案究竟是否需要像图书一样用情报检索语言进行控制;如果需要是否也要像图书那样用国会图书馆主题词表来标引档案。当时,档案界不同意使用情报检索语言的观点占上风。

为了解决这个问题,美国档案工作者协会的国家信息系统特别工作组,专门研究了这个问题,认为:档案的信息系统的建设应该与国家整体的信息系统建设相一致,所以采取了回避争论的方式,制定了美国档案机读目录格式。在美国档案机读目录格式中,规定主题分析块中的 650 字段、651 字段、655 字段、656 字段及 657 字段所使用的主题词必须使用美国国会图书馆主题词表的主题词。虽然,美国档案界在使用美国国会主题词表时感觉不是很满意,其主要意见是,认为国会主题词表中的主题词反映档案主题还不够专指。但是,为了使档案的目录记录进入网络目录数据库,如世界最大的网络目录数据库:美国研究图书馆网络和计算机联机图书馆网络,使档案目录记录能够在这些自动化网络中得到最广泛的利用,美国档案界也就妥协了。正是在这种情况下,美国档案界在进行目录数据库著录时,就不得不使用情报检索语言,从而也把美国档案界推向广泛使用情报检索语言的阶段。

现在美国档案界使用国会主题词表标引档案已经是很普遍的作法了。由于，在美国档案信息高速公路上，所使用的情报检索语言与图书馆界是完全一致的，所以，在形成完整的国家信息系统方面其集成非常度高。

现在，在美国档案界已形成共识：即如果一个信息系统不提供面象主题和内容的主题控制，这就等于自取灭亡。显然，这种对情报检索语言的认识已充分达到了信息是战略资源的高度了。

1.2 加拿大档案界情报检索语言的运用

加拿大档案界情报检索语言的运用，较早的是从不列颠哥伦比亚省的档案信息高速公路建设开始的。本世纪90年初，不列颠哥伦比亚省的档案界开始建设本省的档案计算机联合目录，在开始设计时，该省就把情报检索语言的运用作为数据值的标准来考虑，选用了美国70年代编制的《艺术与设计主题词表》作为控制语言。到了1993年，加拿大档案馆又提出了应该在档案界建立一部统一的综合性档案主题词表，以适应加拿大档案信息高速公路建设的需要。

为什么加拿大也要建立一部统一的综合档案主题词表呢？这是因为加拿大档案界认为：各种形式的档案信息是一种战略资源，应该在其整个生命周期得到有效的管理，所以应该采取必要的方法以便更方便、更准确的查找这些信息，改善决策者的决策，而这种方法就是采用主题词法（叙词法）。为此，加拿大档案馆专门成立了三个工作小组分别研究如下三个课题：检索需求定义；叙词表标准的评价；控制词汇和叙词法。

现在加拿大档案界正在实施的加拿大档案信息高速公路，就特别注重情报检索语言控制的规范化，其原因概出于此。根据加拿大档案界的设计，综合主题词表是各专业表的母表，在信息高速公路上检索时，可以从这一部母表兼容地查到各部门的专业表，从而实现信息检索的无缝联接。

1.3 澳大利亚档案界情报检索语言的运用

澳大利亚档案界情报检索语言的运用是随着澳大利亚档案界信息高速公路的建设——澳大利亚政府信息查找定位服务系统建设而兴起的。为了改善在网络环境下的档案信息检索，澳大利亚档案界提出了基于职能词表的档案管理。职能词表就是主题词表，根据澳大利亚档案界的定义，职能词表与主题词表其间的差别就是：主题词表的词是从文献中选出来的，而职能词表的词是从形成机构的档案中选出来的，这种差别的作用就是使职能词表能够反映机构的事务和活动。

在职能词表的基础上，澳大利亚档案界又提出了自然语言主题词表的概念，认为在万维网上自然语言检索就是显著的例子，所以为了适应自然语言的挑战，应该建立自然语言主题词表。自然语言主题词表由三层构成，第一层是一个由用户用以选词进行检索的综合性词表，第二层是用以反映主题词表控制词汇中的纵向和横向关系，第三层是用以反映自然语言和主题词之间的关系。利用这种词表来提高检索词与标引词之间的匹配程度，实现信息高速公路上从文本或数据库的自动化检索。

在此基础上，澳大利亚档案界还进一步探讨了基于自然语言主题词表的自然语言检索法，如语法分析法、模糊逻辑法以提高检索词与标引词之间的匹配度。

……

从以上世界各国档案界的发展情况可以看出，信息高速公路的建设已经迎来了情报检索语言在档案界发展的又一次机遇。

2 情报检索语言在档案界发展的逻辑探讨

2.1 信息的标准化是情报检索语言在档案界发展的重要前提

在世纪之交,在人类迈向信息社会的时刻,档案信息高速公路的建设理所当然地引起各国档案界的重视,既然要进行信息高速公路建设,标准化是其中不可缺少的组成部分,从某种意义上来说,信息高速公路就是一个多层次、多维的、立体的标准化综合系统。就是由于有了这样一个标准化的信息环境,情报检索语言在这个环境中才得以充分展示其作用和功能。没有这样一个环境,仅有情报检索语言是很难发挥作用的。美国的例子就很好地说明了这一点。

美国国会主题词表编制于1898年,出版于1909年,1988年其结构开始演变为标准化的叙词表结构,这部主题词表是美国最具权威的主题控制词表。但是尽管如此,在本世纪的绝大部分时间里,并没有引起美国档案界的注意。一直到了1985年以后,在美国档案界有了机读目录格式,建立了国家档案信息系统后才使用了该主题词表。这里的一个关键原因就是情报检索语言作为数据值的标准,它必须要有数据内容标准及数据结构标准与其相配套。

情报检索语言,尤其是主题词语言本质上是自动化环境下使用的检索语言,在自动化条件下,其需要有一个标准化的环境,如果这个环境不具备,情报检索语言仅靠其自身的优势要获得档案界的广泛使用就比较困难。而一旦这个环境具备,情报检索语言的使用就可以充分展示其性能。在美国,这个环境就是美国档案机读目录格式的产生,由于有了这一格式,使得美国档案界的档案记录信息就能够进入信息高速高路,同时,这个格式作为数据结构标准,又为档案界使用数据值标准——美国国会主题词表奠定了基础。

这正如1991年张琪玉教授在《情报检索语言在档案工作领域应用的进展》一文中所归纳的,情报检索语言在档案工作领域中迅速地得到推广应用的主要原因是档案工作现代化的需要和文献工作标准化的需要。

2.2 信息的规模化利用是情报检索语言在档案界发展的本质要求

世界各国纷纷建立档案信息高速公路,其本质要求是为了达到对档案信息的规模化利用。所谓信息规模化利用就是以网络数据库为依托,通过信息高速公路以达到信息资源的有效共享,从而实现作为战略资源的档案信息的开发。

在信息社会,作为一个国家的实力强弱的一个很重要的标志就是对信息规模化利用的程度,这是因为作为信息资源,档案的信息价值是智力的外化,这种智力资源一旦被激活(即被得到开发与利用),就可以内化为人们的思想,形成新的创造力,因此,一个国家如果不能对信息资源达到规模化的利用,在信息社会就难以支持这个国家的创新体系。同样,作为人类社会也是如此,如果一个信息时代的文明社会,没有信息资源的规模化利用,那么也就难以支持这个文明社会的创新、进步与发展。这从加拿大的档案界建立信息高速公路的例子就可以说明这一点。

如何达到对档案信息资源的规模化利用呢?这就要求有统一的、高质量的情报检索语言作保证,如在加拿大档案联合目录中,在对一条档案记录进行主题词标引时就规定了从人名、机构名、会议名、地名、普通主题、档案类型、职能和职业等八个方面去提炼主题,而且每一个方面都有一个相当于用于标引的分面公式的子字段列表,所以标引的一致性很高,由于有了分面公式,所以使得进入这个网络数据库的上百个档案机构在主题词的标引方面大大降低了标引难度,提高了标引的方

便性和检索的一致性，十分有利于档案信息资源的规模化利用。

2.3 图书馆界在创制和使用情报检索语言方面的成功经验是情报检索语言在档案界发展的重要借鉴

应该说，情报检索语言在档案界发展是以图书馆界的经验为基础的，如果没有图书馆界在创制和使用情报检索语言方面的成功经验，也就很难有情报检索语言在档案界的发展。这从澳大利亚档案界建立档案信息高速公路的事例中就可以看出这一点。澳大利亚档案界在研究自然语言检索时就充分借鉴了美国芝加哥大学和美国卡纳基·梅隆大学的经验。情报检索语言在档案界发展的这一动向充分证明了张琪玉教授在八年前的预言：档案界对情报检索语言的研究和应用，其方向是坚定的、其范围将继续扩大，其速度将会更快。

3 我国档案界情报语言运用的现状与发展策略

3.1 我国档案界情报语言运用的现状

首先，我国情报检索语言的运用应该说有了一个很好基础，1987 年我国就编制了全国统一的综合性分类法——《中国档案分类法》，1988 年又编制了全国统一的综合性主题词表——《中国档案主题词表》。从目前国际范围内来看，在档案检索语言的创制方面，我国的确是领先的。但是从应用现状来看，除《中国档案分类法》使用较多外，《中国档案主题词表》使用范围却相当有限。

其次，我国的档案信息资源的利用远没有达到规模化利用的阶段，据了解全国三百多个公共档案馆的接待利用率，为平均每天接待一人次。所以，档案馆反映标引主题词没有用，用户查档案不用主题词。

再次，我国档案界目前在网络环境下的标准不系统，不成体系，形不成档案信息高速公路，由此阻碍了情报检索语言的更广泛地运用。

第四，我国档案界虽然也进行机读词表、自动标引等情报检索语言的研究，但是，由于用户面狭窄，转化为实际应用的成果就很少，从而也就限止了档案界在更深层次上的情报检索语言研究。

此外，从近年研究的文献看，我国档案界关于情报检索语言的研究文章的数量也很少。

3.2 我国档案界情报语言运用发展的策略

档案信息是我国重要的战略资源，因此，在信息这层面上，为达到我国档案信息资源的规模化利用，必须建立我国自己的档案信息高速公路，通过档案信息高速公路的建立，以深化我国档案界情报检索语言的运用，从而更有效地开发和利用档案信息的战略资源。为此，我们应该采取如下策略：

1. 建立与情报检索语言相配套的标准化体系，以改善情报检索语言的运用环境，使情报检索语言全面地、真正地用起来。

2. 明确档案信息规模化利用的基础是网络数据库，因此，及早规划我国档案网络数据库的建设，改变目前我国档案馆数据库建设无标准、分散多头、而没有规模效益的现象。

3. 加强国际档案界有关情报检索语言的运用情况的了解，正确分析与制定我国档案界情报检索语言的发展规划。

4. 更有效地与图书馆界合作，努力借鉴图书馆界在创制与研究情报检索语言方面的成功经验，

以最大的费效比，加速情报检索语言在我国档案界的运用。

参考文献：

[1]张琪玉.情报检索语言在档案工作领域应用的进展.文献工作研究，1991(6)
[2]Australian Governments Interactive Functions Thesaurus, 1998
[3]CA . Thesaurus as a tool for the management of government information, 1996
[4]U.S. Library of Congress Subject Headings ,1994

开创情报语言学的新天地
——论张琪玉教授对情报语言学的新贡献

中山大学信息管理系　曹树金
中山大学图书馆　罗春荣

[文摘]论述张琪玉教授近些年来在自然语言检索、情报检索语言和标引自动化、未来理想检索语言模式等方面的研究成果,说明他对我国乃至世界情报语言学的巨大贡献。

[关键词]情报语言学　研究领域　贡献　张琪玉

1　导言

70年代末,80年代初,张琪玉教授以保证情报检索系统的检索效率为中心,开拓性地对各种情报检索语言进行系统性、整体化的研究,著就《情报检索语言》一书,标志着情报语言学的诞生,张教授也就成为情报语言学的当之无愧开山鼻祖。正如吴建中所言,"80年代初,张教授针对当时人们把注意力集中在体系分类法的思想性方面,使图书分类研究陷入所谓'三性(思想性、科学性、实用性)'怪圈问题时指出:'要改变研究方向,把研究的重点转移到如何提高情报检索语言的检索效率方面来',并独辟新路,开创'情报语言学',提出从更高的层次对分类法、主题法等各种检索方法进行统一研究。张教授这一开拓性的研究,对我国情报检索语言领域的理论与实践起到了积极的导向和推动作用。"[1]范并思进一步指出,"与图书馆现代化相比,检索语言研究有更多的创新意义。图书馆现代化研究是学习西方建立起来的领域,面检索语言研究却基本属于中国人自己创立的领域。在张琪玉以前,国外还没有人能在"情报检索语言"的书名下将检索语言理论的内容讲得如此系统面精密。"[2]

此后,张琪玉教授更加将全部身心倾注在情报语言学的研究、实践和教育上,不分寒暑,从无假日暇时,孜孜不倦地探索和耕耘。尤其是最近几年,以逾花甲之龄,张教授年均发表情报语言学论文10余篇,为情报语言学的深化和拓展做出了新的巨大贡献。至今,他已经发表论文译文160余篇,论文被各种文集等转载约60篇次;编著、主编、参编的专著有20多种。因在编制《中国图书馆图书分类法》中作出较大贡献面获1985年国家科学技术进步奖一等奖,因在编制《中国分类主题词表》中作出贡献而获1996年国家优秀科技信息成果二等奖。

其中,最有代表性的新贡献反映在《情报检索语言》不断增加的内容上。1987年将《情报检索语言》增订更版为《情报语言学基础》,主要增加"第一章 绪论——情报语言学 ·般问题"、"第九章 情报检索语言基本方法"和"第十一章 情报检索计算机化与情报语言学的发展",使情报语言学的内容更加充实;1997年出版的《情报语言学基础》增订二版,又增补"分类法主题法一体化检索语言"和"自然语言在情报检索中的应用"两章,使情报语言学上升到一个新的高度。这些不断的增

补,既是国内外情报语言学领域新进展的及时反映,更是张琪玉教授不断取得的创造性成果的凝聚。1980－1997 年期间,被 288 位作者的专著和论文共引用了约 450 次。作为教材,在图书馆学、情报学的本科、函授和电大等教学中被广泛采用,估计使用该教材学习过这门课程的不少于 25000 人。从其成果的广泛传播,不难看出张琪玉教授的情报语言学研究成果的先进性、实用性及其巨大影响。

如果我们全面总结、认真学习和借鉴张琪玉教授的治学精神、研究方法和丰硕成果,相信能够促进我国情报语言学更快、更好地发展。有鉴于此,本文将着重讨论张琪玉教授近些年来在自然语言检索、情报检索语言和标引自动化、未来检索语言模式等方面的研究。

2 对自然语言检索的研究

从 80 年代开始,自然语言检索(自然语言应用于情报检索)成为国外情报检索和自然语言处理领域的共同研究热点。这从 1981－1995 年间,ERIC,INSPEC,Compendex Plus,LISA 和 Information Science Abstracts 五个数据库所收录的"全文检索"的文献数分别是 133、471、40、525 和 107,就可见一斑[3]。我国对自然语言检索的研究相对较为薄弱。但是,张教授是国内较早关注自然语言检索的学者之一,至今,他对自然语言检索的研究最为全面、系统。实际上,近些年,张教授已经将自然语言检索作为自己的研究重点,其主要研究成果已集中反映在《情报语言学基础》增订二版的第十二章中。[4]

2.1 对自然语言检索研究的呼吁

早在 1983 年张教授就指出"我们的情报检索语言理论研究工作必须开辟新路。……关于扩大研究范围,似乎也应把对自然语言检索法的研究包括进去。电子计算机检索技术的发展,为自然语言检索法开拓了光明的前景。自然语言检索法最终是否能取代以及何时能取代人工语言即情报检索语言,目前下结论还为时过早。但是,人工语言和自然语言相结合,既是严密的人工语言,又尽量利用自然语言作为辅助,无疑有助于检索效率的提高,因而是情报检索语言发展的正确方向。"[5]

不久之前,他回顾了我国的自然语言检索研究进展[6],鉴于"目前所能见到的论述自然语言检索的文章,大多只是将自然语言与人工语言作笼统的比较,深入考察自然语言性能者不多见。"他呼吁"情报语言学研究者应当积极参与自然语言检索的研究,当前亟需从情报语言学角度深入研究自然语言检索方法,把情报语言学的原理和方法引进自然语言检索的研究。"[7]

2.2 自然语言检索与情报检索语言关系的辩证观点

根据情报检索的基本要求以及自然语言检索和情报检索语言的各自优缺点,张教授一方面肯定,"在计算机检索越来越发展的条件下,自然语言具有不可阻挡的发展前途。特别是在互联网络的检索环境中,它将成为一种必然的优先选择。"另一方面,并不赞同自然语言将替代情报检索语言或最终将替代情报检索语言,情报检索语言或情报检索语言研究已经过时的论点。他认为:自然语言检索系统与情报检索语言检索系统并不是绝然对立的,它们各有长处和短处,可以并行发展,可以互相结合,互相补充。虽然自然语言的初级应用(如关键词检索、文本匹配查找)迅速扩大,但一些著名的检索工具和数据库并没有放弃人工语言的迹象。自然语言要全面胜过人工语言是不可能的,除非它引进许多情报检索语言的原理和方法,而不是单纯的自然语言。对于高要求的情报检索来说,控制是绝对必要的。而"对检索过程进行控制",正是情报语言学的精髓。我们应当把情报语

言学的理论成果运用到自然语言方面去。为了提高自然语言的检索效率,需要采取后控制措施,特别是采用后控制词表。将来有可能使用一些人工语言与自然语言结合的、带有自然语言换词、换号功能的分类表和词表。语义关联对情报检索是绝对必要的,但既可以在先控制系统中,也可以在后控制系统中。[8]

至于自然语言或情报检索语言的未来,他认为将是自然语言的情报检索语言化或情报检索语言的自然语言化。未来的检索语言将既是情报语言,但不是现今的情报检索语言模式;它既是自然语言,但也不是现今的自然语言检索模式;它既具有自然语言的优点而优于现今的情报检索语言,又具有情报检索语言的优点而优于现今的自然语言。

2.3 对自然语言检索中影响检索效率的各种因素的深入分析

张教授对自然语言检索的研究,秉承他一贯以检索效率为核心的传统。因此,他具体论述了影响自然语言检索效率的主要因素:检索所依据的文本的类型(文献题名、文献中的小标题和章节名、文献的摘要和正文);检索用语的专指度;在文本的不同范围(句段、节、篇)内进行组配检索;文本用词的不规范性;不同标引方法(不标引、自动抽词标引、人机结合抽词标引、自动赋词标引和自由标引);对自然语言进行词表控制的程度。在此基础上,提出了改善自然语言检索效率的基本原则。[9]

通过调查国内外的相关文献,我们发现,张教授对自然语言检索中影响检索效率的各种因素的分析,最为深人全面。这对正确认识自然语言检索存在的问题,探寻有效的改进方法有很大的启发。

2.4 自然语言在情报检索中应用方式的全面总结及有待解决难题的明确

为了纠正不少人对自然语言应用于情报检索的片面理解,开阔人们的视野,张教授全面总结了自然语言在情报检索中应用的方式。他指出,关键词法;文本检索;单汉字检索;以自然语言作为自由词进行补充标引,与情报检索语言结合使用;以自然语言作为人口词(接口),利用计算机的换词功能,辅助情报检索语言;自动赋检索词和自动赋分类号;自动分类(自动聚类法);自由标引等都是自然语言应用于情报检索的方式。[10]

张教授明确指出,自然语言在情报检索中的应用,面临着两难题,一是如何从自然语言文本中抽出最能准确、充分地表这文献有价值内容的词,以及这些词与检索课题有效匹配的问题。如果去追求百分之百的自动化,至少在短期内是无希望解决的(当然,自然语言自动处理现有的一些中间成果还是有实用价值的)。如果采用人机结合的方法,则可以较为容易一些;二是克服自然语言由于不规范和缺乏语义关联性而对检索不利的问题。此外,对中文来说还有一个汉语分词的问题。10 多年来,我国进行了大量的汉语分词技术研究,提出了许多分词方案,总的来说,在这方面已取得了很大进展,现在距离解决汉语自动分词的问题已为期不远。但这个问题的解决,只是达到了拼写文字国家的起点水平,拼写文字中未解决的上述两个问题仍有待我们去解决[8]。这实际上是强调了我们研究自然语言检索应该把握的方向。综观国外的自然语言检索研究,我们发现,他们正是围绕这两个难题而展开。

2.5 对后控制词表的系统研究

张教授对后控制词表非常重视,他在多篇论文中强调后控制词表是弥补自然语言缺陷的主要措施,不仅自己发表《论后控制词表》一文[11],而且先后指导两位研究生以"后控制词表检索系统研

究”和“全文检索系统后控制技术研究”作为硕士论文题目。

在《论后控制词表》中，系统论述了后控制词表的控制机理、控制程度、后控制词表编制上的特点、后控制词表的各种编制方式及其在控制上的差别，提出并用实例说明了一种“分类词表+字顺/轮排表”的后控制词表模式，阐述了利用后控制词表检索文献的各种方法。

2.6 关于自由标引的论述

自由标引是情报检索中利用自然语言的一种简易方式，我国不少单位在专题数据库建设中采用这种不依据词表进行主题标引的方式。但是，专门论述自由标引的文献很少。张教授的《论自由标引》[12]一文，全面阐述了自由标引的概念、自由标引的优点及适用范围（主要适用于报纸文献、期刊文献的大型篇名数据库等标引）、自由标引种类、自由标引基本方法、自由标引要点、自由标引系统配备后控制词表的必要性，有益于人们在理论上准确认识自由标引，在实践中正确采用自由标引。

2.7 对汉语关键词法的探讨和试验

在《情报检索语言》一书中，张教授就将关键词法作为一种“准情报检索语言”予以重视[13]。后来，针对关键词法应用于汉语的特殊问题，不仅进行理论探讨，而且亲自编写程序，利用计算机进行关键词索引的编制试验，先后发表了多篇论文。他不仅自己亲自研究，而且还指导多位研究生研制“汉语题内关键词索引与后控制词表系统”，在《文献工作研究》1994 年第 3－4 期发表 6 篇系列报告。

在《汉语关键词法探讨》一文中[14]，论述了关键词法在我国的实用价值，尤其是对中文报刊文献的标引，关键词法不失为一种好的选择。分析了汉语文献题名进行自动抽词的进展与问题，指出不可能仅凭非关键词表用排除法抽词，而必须用能辨识实词和虚词的抽词词典，但是这种词典编制起来很困难；由计算机自动从汉语句子中抽词（中文自动切分或自动分词）取得了很大进展，创造了很多切分方法，但是中文自动分词是不可能彻底解决的。因此，人机结合抽词的技术路线（指人工抽词后由计算机整理的方式，以及由计算机抽词并辅以人工干预（判别）的方式）目前仍有现实意义，是一种合理的过渡办法，既可达到提早实用的目的，又可为接近完全的自动抽词创造条件。文中讨论了人工抽词－计算机整理方式和计算机抽词－人工干预方式的具体问题，比较了人工抽词、人工干预和自动抽词三种方式的各自优缺点，最后阐明抽词词典、规范词典和后控词表的编纂对发展汉语关键词法的重要意义。

在《人－机结合的题内关键词索引可回避汉语分词难题》中[15]，他指出汉语题内关键词索引可以“含糊抽词”，实际上是“最长抽词”与“词素轮排”的结合。而且还可以对题内关键词索引加以改进，例如，在索引条目中增加著者项几乎取消题录部分，可节省很多篇幅；便于对文献进行甄别；著者也参与轮排，更增加了著者索引的功能。

《汉语关键词索引的一种编制方法》[16]和《汉语关键词索引的另一种编制方法》[17]介绍了张教授将 WPS 与 dBASE 结合编制题内关键词索引的两种简易方法及实现程序。

从文本（主要是题名和文摘）中自动抽取关键词是较易实现、尚可接受的一种自动标引方案。近二十年来，我国学者对汉语分词（从文本中自动抽出关键词）技术作了许多研究，提出了不少分词方案，但见于实际使用者不多。张教授认为，原因并不是这些方案都经不起实践考验，而主要是缺乏抽词词典或缺乏完善的抽词词典。因为只有抽词软件而无抽词词典，是不能建立自动抽词标引系统的。目前缺乏抽词词典，已成为限制抽词标引技术推广应用的瓶颈。我们图书情报工作者应

当积极地承担编制抽词词典的任务。而且，可将编制抽词词典与建立实际需要的数据库相结合。[18]

此外，张教授对汉语自动分词、单汉字检索等也进行了总结研究。

3 对情报检索语言和标引自动化的研究

在1987年出版的《情报语言学基础》最后一章“情报检索计算机化与情报语言学的发展”中，张教授根据自己的考察，综述了计算机技术的发展对情报语言学的重大影响[19]，即：50年代开始的情报检索计算机化，促进了情报检索语言的创新和改造，使词表、分类表向机编化和机读化方向发展，使文献标引和索引编制走向自动化，使自然语言检索得以实现，使多种语言的结合使用成为可能，使检索方法有了很大的进步，并正在使情报检索语言的应用范围扩大。同时，情报检索计算机化对情报检索语言研究提出了许多新课题，并提供了许多新方法和新条件。总之，情报检索计算机化对情报语言学的发展产生了极为深刻的影响。因此，他认为，在计算机检索正在逐步发展到互联网阶段的新形势下，情报语言学研究的基本课题应是如何使情报检索语言适应新的检索环境。例如分类表、词表如何在新的检索环境中继续发挥作用，是一个有待探索的课题。[7]

3.1 自然语言与人工语言对应转换是情报检索语言走向自动化之路

在实现情报检索语言和文献标引自动化的道路选择上，张教授明显坚持务实的风格，也就是注重研究当前切实可行的技术方法。基于他对概念或标识自动转换技术可广泛应用于文献标引和情报检索的认识[20]，张教授论证了自然语言与人工语言对应转换是情报检索语言走向自动化之路的观点[21]。

在计算机检索系统条件下，对文献的标引模式可以有19种，这19种都属自动化范围的模式可以归纳为(1)人工标引，自动检索；(2)人机结合标引，自动检索；(3)自动标引，略加人工辅助，自动检索；(4)自动标引，自动检索。总之，在计算机检索系统检索(查找)过程都是在程序控制下自动进行的，但标引过程却可以是多种多样的，有许多种标引模式。

自动或半自动标引(无论是自动抽词或自动赋词还是自动分类)的处理对象都是文献本身的用词(即自然语言)，所以，自动抽词和自动转换是自动标引的主要内容。

自动转换必须以自然语言与人工语言的对应为前提。通过对应表将自然语言转成人工语言。所以，把分类表和词表改造成自然语言与人工语言的对应表，是情报检索语言走向自动化的必由之路。把词表改造成自动赋词适用的词表，其实质就是增加大量自然语言词作为入口词。体系分类表改造成自动赋号适用的分类词表——词—分类号双向对应表，较为复杂，而且还要制定自动分类规则，但实践已证明是可行的。

《中国分类主题词表》的机读版，无论从增加自动标引功能考虑，还是从增强其机辅标引功能和在计算机系统中更充分地发挥检索功能考虑，都有必要作这种改造。若要使其具有自动赋词功能，只要将其第二表改造成为自然语言词－主题词－分类号对应表即可，若要使其具有选定主要分类号的功能，则还需对其第一表进行改造，但不必改动原有分类号。

3.2 《中图法》的自动化

张教授认为，《中图法》应该尽快自动化，新一版《中图法》就应是一部多功能的电子版文献分类法。[22]

它应是在《中国分类主题词表》的基础上编制的，并要把自然语言对应进去，具有自动抽词、自动赋号、自动赋词功能，可提供分类号、主题词、关键词三种检索途径；优化《中图法》的检索功能，可更多地放在电子版中。如《中图法》的分面化改造，就可以在电子版中实现（由于自然语言与体系分类表的对应转换要通过“关键词→组配分类表号类号→体系分类表类号”的间接对应方式），但排架分类号可原封不动，即利用对应转换技术，根据类目内容范围划分规则，将电子版中组配分类号标引的概念仍对应到印刷版中原来应入的分类号中去就可以了。

使《中图法》电子版具有自动赋号和自动赋词的功能，还必须编自动抽词词典。但这种抽词词典可以说是为电子版增加自然语言与人工语言对应转换功能的副产品，只需配备一个汉语自动分词软件就行了。

3.3 易于实现的自动分类方法

为了探索比自动聚类更为简单、可靠的自动分类方法，张教授仔细研究了基于分类表结构原理和通常的类目内容范围划分规则的自动赋号法。[23]

自动分类必须以自动抽词为前提。中文的自动抽词必须借助于用词词典，或用词词典与非用词词典并用。当然，开始的时候也可能采用半自动方式抽词。

以体系分类法为基础的文献自动分类需要一种自动分类用分类词表。这种分类词表是词－分类号双向对应表，分为分类号－词对应表和词—分类号对应表两部分。如果使用《中图法》，需将《中图法》分类表改造成分面分类表，把词对应到相应的分面中，但原有类目内容和分类号不需要改变。

词－分类号对应表是倒转过来编成的，按词的字顺排列。这个表用于对文献自动分类标引。

在自动标引过程中，将从文献题名中自动抽出的词通过与词—分类号对应表核对，赋予《中图法》的分类号，建立分类号索引，提供分类检索途径。同一题名中的词因为分属于不同的分面，其分类号也就有多个。词仍应保留，建立词索引，提供主题检索途径。

如果一篇文献获得的多个分类号中，需要确定一个主要分类号，可根据通行的分类规则，将各个分面定出一个优先次序。一篇文献所获得的几个分类号中，哪个分类号所属分面在别的分类号所属分面之前，就确定那个分类号为主要分类号。如果遇有两个分类号所属分面并列，则两个分类号均为主要分类号。

如果以分面分类法为基础的文献自动分类，其所用分类词表也分为两部分。但是，分类号－词对应表部分的编制极为简单，只要把从文献题名中自动抽出的词对应到分面分类表相应的类目下即可。词－分类号对应表也只要将分类号－词对应表倒转来就成了。

用这种自动分类方法建立的检索系统，实际上是分类法主题法一体化的检索系统，可以进行分类（组配）检索、主题（组配）检索、分类与主题组配检索。

4 对未来理想情报检索语言的探索

张教授非常重视运用他自己总结出来的情报语言特有研究方法，去把握情报检索语言和情报语言学的发展方向，挖掘研究课题，解决新问题。他之所以能在情报语言学研究中取得丰硕成果，在很大程度上是因为能够娴熟地运用合适的研究方法。其中，主要的研究方法是结构功能分析法和理想语言设计法。

为充分发挥情报语言学理论对实践的导向作用，超前地预见情报检索语言的未来，十多年来，

张教授一直遐想着运用理想语言设计法和结构功能分析法，创制一种完全新颖的、能够满足人们对情报检索语言的绝大部分要求的情报检索语言结构模式。

4.1 未来理想情报检索语言的要求

张教授给这种理想的情报检索语言设想的结构要求是：学科聚类系统与事物聚类系统的结合（事物聚类也应当有系统性，字顺序列可作为进入事物聚类系统的手段），先组式语言与后组式语言的结合，体系分类法与组配分类法的结合，人工语言与自然语言的结合，号码标识与语言标识的结合，系统序列与字顺序列的结合，不变概念代码与可变概念体系的结合。

他给这种情报检索语言提出的功能（性能）要求是：分类法与主题法彻底一体化；充分发挥情报检索语言对知识进行系统组织和对自然语言进行规范控制的功能；用户十分方便地进行标引和检索；概念可不断增补及概念的代表词可进行更换；用户区别不出是自然语言还是人工语言，其实是由严密的人工语言控制；修订不受已标引文献所牵制，故分类体系可逐步完善；并可以挂接英文索引、分子式索引；以及可用于机助标引等。

4.2 未来理想情报检索语言的多样性“结合”

张教授诊断，构成这种理想情报检索语言性能的原理和方法都已存在，难题在于找到它们的结合方案。经过长期的思索，他找到了达到上述要求的方案。1997 年，张教授发表《学科－事物概念组配型检索语言——关于情报检索语言的遐想与求索》[24]和《探索 21 世纪的情报检索语言》[25]两篇论文，具体描述了他所设计的理想情报检索语言的原理和方法，他将这种模式的情报检索语言命名为“学科－事物概念组配型检索语言”。

他论述了理想情报检索语言应是学科－事物概念组配型检索语言。它由学科分类系统面和事物分类系统面构成，两个面可以互相组配。当按学科聚类时，藉助于事物及其部分面进行系统复分；当按事物聚类时，藉助于学科及其问题面进行系统复分。这种情报检索语言既有很强的系统性，又有很大的灵活性。它含有分面组配分类法的成分，但不同于分面组配分类法；它也含有分面叙词法的成分，但也不同于分面叙词法。

它设计的这种检索语言在本质属性上实现了多样性的“结合”：

(1)学科聚类系统与事物聚类系统的结合。它能够同时提供学科分类系统和事物分类系统，两者以“学科－面－点：事物－面－点”或“事物－面－点：学科－面－点”的形式构成先组式标引句，既可按学科完全集中文献，又可按事物完全集中文献。

(2)先组式语言与后组式语言的结合，体系分类法与组配分类法的结合。这种检索语言按其本质是一种后组式语言，但当它采用上述形式的标引句并将其系统排序时，就具有了先组式语言的性能。同时，在计算机检索系统中，采用上述形式的标引句，并不妨碍进行自由组配检索，因而它并未丧失后组式语言的性能；它既是体系分类法，又是组配分类法。

(3)人工语言与自然语言的结合。在这种检索语言中，检索系统内部用于标引文献的，既不是分类号，也不是概念词或自然语言词，而是用概念代码。分类号、概念词和自然语言词都与概念代码对应，作为概念代码的外部形式，三者在标引和检索中可任意使用，通过计算机与概念代码自动转换。自然语言可大量使用，任意增补，但在系统内部是受到控制的。

(4)号码标识与语词标识的结合，系统序列与字顺序列的结合。在这种检索语言中，既使用号码标识（分类）号，也使用语词标识（概念词和自然语言词），两者完全对应，具有等价关系。既然号码标识与语词标识并存，就必定可以同时提供系统序列和字顺序列，也即具有分类检索语言和主题

检索语言双重特征。

(5)不变概念代码与可变概念体系的结合。设置不变的概念代码是这种检索语言的特异之处。概念代码是学科和事物概念登记号(顺序号),它固定不变,始终与某一概念相对应,文献实际上是用这种代码来标引的。这样,不变的概念代码是这种语言的主体,而分类号、概念词(代表某个概念的正式词)、自然语言词都与概念代码相对应,仅仅是概念代码的索引而已。概念代码仅在系统内部使用,标引人员和检索人员使用的仍是分类号和语词。采取这种措施,就使概念分类体系具有了极大的灵活性,分类体系的改变与对文献已作的标引无关,概念词和自然语言词的字面上的改变也与对文献已作的标引无关。

4.3 未来理想情报检索语言的实现方法

这种学科-事物概念组配型检索语言的主要实现方法是"分面分析+概念代码+概念对应转换+数据库技术"。

它对概念进行分面化处理,将分面分为学科面部分和事物面部分,每一部分再分为第一层的分面(学科或事物)和第二层的分面(学科的问题或事物的部分)。

该检索语言由多种文档组成。其中,主文档设置下列字段:①概念代码(分2个子字段);②分类号(分4个子字段);③概念词或自然语言词;④子文档区分号;⑤临时分类;⑥轮排标志;⑦参照和注释。由主文档生成概念代码索引文档、概念分类索引文档、概念词索引文档、语词轮排索引文档。概念分类索引文档、概念词索引文档、语词轮排索引文档都可用于联机标引,也可作为检索途径转入供检索用的文档。

文献标引数据文档包括下列字段:①概念代码;②概念代码;③概念代码;④概念代码;⑤概念代码;⑥概念代码;⑦文献类型复分;⑧文献号。将文献标引数据文档利用概念代码索引文档转换成供检索用的文档。检索用文档包括下列字段:①学科面分类号(3个字段);②事物面分类号(3个字段);③文献类型复分;④文献号。供检索用的文档可根据需要变换字段次序,从而产生不同的分类体系,达到不同的聚类效果。将概念分类索引文档挂接于供检索用的文档,在检索界面上可随分类号的变动动态地显示对应的概念词。

对词典的各种修改都在主文档中进行,修改后重新生成各种索引文档和供检索用的文档。

这种情报检索语言模式,在全面继承现有情报检索语言成果的基础上,为下个世纪的情报检索语言设计提供了一种新的思路,这在国内国际都是一个创举!即使是美国国家医学图书馆主持开发10余年的UMLS(一体化医学语言系统)的超级叙词表(Metathesaurus),虽然是以概念为中心,将每个概念与多个术语及更多的词串对应,给每个概念、术语、词串唯一的编号,为概念名称编有指向概念、术语、词串编号的单词索引、正规单词索引、正规词串索引;虽然将众多来源的叙词、标题词、分类法的类名和大量自然语言词汇整合在一起,但是,作为一个整体,它主要是为后控制使用,它没有建立分类体系来实现分类主题一体化,概念代码不用于标引[26]。因此,没有达到张教授所设计的理想情报检索语言模式的水平。

欣逢恩师70寿辰,弟子不才,撰此学习心得,是为贺。恭祝恩师健康长寿,为情报语言学的发展做出更大贡献。

参考文献:

[1]吴建中.21世纪图书馆展望——访谈录.上海科学技术文献出版社,1996
[2]范并思.论图书馆学学科前沿的转移.图书馆,1993(4)

[3]Sievert, Mary Ellen C. Full－text Information Retrieval : Introduction . Journal of the American Society for Information Science , 1996,47(4)
[4]张琪玉.情报语言学基础(增订二版).武汉大学出版社,1997
[5]张琪玉.论情报检索语言的研究、创制与普及图书情报知识.1983(4)
[6]张琪玉.自然语言检索研究进展.见:马费成主编.知识信息管理研究进展.武汉大学出版社,1998
[7]张琪玉.世纪之交中国情报语言学发展之路.图书馆杂志,1997,增刊
[8]张琪玉.情报检索语言的发展趋势(与吴建中的对话).图书馆杂志.1996(4)
[9]张琪玉.自然语言检索中各种因素对检索效率的影响.情报理论与实践,1997(5)
[10]张琪玉.自然语言在情报检索中的应用.情报理论与实践,1996(3)
[11]张琪玉.论后控制词表.图书情报工作,1994(1)
[12]张琪玉.论自由标引.图书馆学刊,1995(5)
[13]张琪玉.情报检索语言.武汉大学出版社,1983
[14]张琪玉.汉语关键词法探讨.图书馆杂志,1993(4)
[15]张琪玉.人－机结合的题内关键词索引可回避汉语分词难题.图书馆杂志,1993(4)
[16]张琪玉.汉语关键词索引的一种编制方法.图书馆理论与实践,1998(1)
[17]张琪玉.汉语关键词索引的另一种编制方法.图书馆理论与实践,1998(4)
[18]张琪玉.缺乏抽词词典是自动抽词标引难以普及的主要原因.图书与情报,1998(2)
[19]张琪玉.情报语言学基础.武汉大学出版社,1987
[20]张琪玉.概念或标识自动转换技术的应用.图书馆杂志,1998(6)
[21]张琪玉.自然语言与人工语言对应转换——情报检索语言走向自动化之路.中国图书馆学报,1996(1)
[22]张琪玉.情报检索语言走向自动化之路与《中图法》发展新目标.北京图书馆馆刊,1996(4)
[23]张琪玉.分类法主题法一体化自动标引系统的基本原理和方法.图书馆论坛,1995(6)
[24]张琪玉.学科－事物概念组配型检索语言——关于情报检索语言的遐想与求索.图书馆杂志,1997(2)
[25]张琪玉.探索21世纪的情报检索语言.北京大学学报(信息管理系建系五十周年专刊),1997
[26]U. S. Department of Health and Human Services et al. UMLS Knowledge Sources. National Library of Medicine, 1999

张琪玉情报语言学学术思想简评

空军政治学院信息管理系　傅　亮

[摘要]本文简要评述了张琪玉教授的情报语言学学术思想,包括情报检索语言研究方向的转变,系统研究方法的创建,研究领域的拓展,学科体系的构建以及新型情报检索语言的创制与对未来情报语言学的探索。

[关键词]情报语言学　情报检索语言　张琪玉　学术思想　简评

1　引言

张琪玉教授是中国当代著名的图书馆学家,情报语言学的开拓者,长期致力于图书馆工作、学术研究和教学活动。其主要研究领域为情报语言学。现任空军政治学院信息管理系教授,中国图书馆学会理事、上海市图书馆学会副理事长、中国索引学会副理事长。曾任武汉大学图书情报学院教授、图书馆学情报学研究所所长、空军政治学院图书档案系主任等职。他所编著、主编、参编的专著有20多种,发表论文、译文约160篇。代表作为《情报检索语言》(增补修订本改名为《情报语言学基础》)。曾因在编制《中国图书馆图书分类法》工作中作出较大贡献于1985年获国家科学技术进步奖一等奖;因在编制《中国分类主题词表》工作中作出的贡献于1996年获国家优秀科技信息成果奖二等奖。1991年7月起享受国务院颁发的政府特殊津贴。其简传载入《中国大百科全书》(图书馆学情报学档案学卷)、《世界名人录》(中国卷)等。

张琪玉教授在长达40多年,特别是近20年的学术活动中,致力于情报检索语言研究。他开拓了情报语言学的研究领域,并在长期的研究中形成了丰富的情报语言学学术思想。本文仅是根据自己的学习体会,就几个主要方面作一评述。

2　明确了提高情报检索效率是情报检索语言的研究方向

张琪玉教授认为:作为表达文献主题概念和检索课题概念的语言工具,其结构情况和编制质量以及对它的使用是否正确,是影响情报检索系统效率的主要因素,故如何提高检索效率,是情报语言学研究的根本目的和核心问题。他主张对分类检索语言、主题检索语言和其他情报检索语言以及自然语言在情报检索中的应用问题进行统一研究,以探索它们影响检索效率的共同规律和有效的改进途径。这一学术思想的提出,对我国情报语言学摆脱对体系分类法"三性"(思想性、科学性和实用性)研究的局限,明确情报语言学理论研究的核心和方向,具有极其重要的意义。正如吴建中博士所评价的:"张教授这一开拓性的研究,对我国情报检索语言领域的理论与实践起到了积极的导向和推动作用"。[1]

据曾蕾调查分析,建国后到80年代初,我国图书情报界对情报检索语言的理论研究大多集中

于对体系分类法“三性”及其间关系、部类设置、大类序列、以及科学分类与图书分类的关系等问题上。虽然研究重心向图书分类法的基本性质和指导思想的倾斜,在建国当初,对编制适合新中国国情的分类法具有特定的历史意义,但随着时间的推移和研究的深入,“对究竟如何改进体系分类法、更好体现‘三性’却深入不下去”[2]。当时,只有极少数文章谈及检索效率,而对旧分类法和国外分类法进行批判的文章却很多,总体的研究忽视了情报检索语言作为情报检索语言工具的基本职能。面对这一情况,张琪玉教授以其扎实的知识功底和敏锐的洞察力,适时地提出“要改变研究方向,把研究的重点转移到如何提高情报检索语言的检索效率方面来”[3],使中国情报语言学走上了更为科学的研究道路。他确立了以检全率和检准率为核心的“全、准、快、便、省”的检索效率体系,明确指出情报检索系统效率的高低,在很大程度上取决于所采用的情报检索语言的质量以及对它的使用是否正确。

张教授的这一思想始终贯穿于他的情报检索语言研究之中,尤其反映在其专著《情报检索语言》(后增订本改名为《情报语言学基础》)中,书中详细分析了情报检索语言本身影响检索效率的多种因素,分别对各种情报检索语言从检索效率的角度进行剖析和对比研究,分析其优势和局限并提出改进建议。不仅如此,当面对情报语言学研究的新兴领域——自然语言检索时,他依然遵循着这一研究思路,提出了独到见解,这点可从其发表在《情报理论与实践》1997 年第 5 期上的论文《自然语言检索中各种因素对检索效率的影响》中见诸一斑。

3 创建了系统的情报语言学研究方法

张琪玉教授对情报检索语言的研究之所以能获得成效,一方面是由于其对情报检索语言研究指导思想的转变,另一方面是由于采用了适合情报检索语言这个研究对象的研究方法。具体来讲有:⑴结构功能分析法;⑵历史演进研究法;⑶调查整理法;⑷比较研究法;⑸归纳法和演绎法;⑹原理或方法的移植法;⑺理想语言设计法(设计——编制——使用——优化);⑻现用语言改进法;⑼数学和统计方法;⑽实验法;⑾计算机方法,等等。其中,结构功能分析法是所有其它方法的基础,而其它方法都是从结构功能分析法中衍生出来的。

将系统论的结构功能分析法运用在情报语言学的研究中是张教授的首创。大量研究事实证明了该方法是最适合这门学科研究工作的。他精辟地指出:“情报检索语言功能的完善,都是通过其结构的改进和增加来达到的。要弄清某种结构产生什么功能,或某种功能是由哪些种结构产生的,就必须对各种情报检索语言进行解剖分析,要弄清产生同一种功能的不同结构孰优孰劣,以及某种结构的适用范围和条件,就必须将各种结构进行比较研究。”[4]在具体研究过程中,他也是这样做的。他从宏观和微观的角度对各种情报检索语言进行了剖析,发现无论是分类表还是主题词表,其宏观结构的多少决定了语言可实现功能的多少,而微观结构的性能则决定着某种功能的强弱。此外,他还意识到功能的实现并不是靠各种结构的简单相加来完成的,因此要提高情报检索语言的检索效率就必须从宏观结构的构成和微观结构的质量入手进行调节、控制以及创新。张教授曾在其著作和文章中多次谈及和运用了结构功能分析法,《情报检索语言方法综述》一文更是全面详细地阐述了该方法的基本原理并运用它具体研究分析了情报检索语言为发挥其四大基本职能所使用的各种宏、微观结构。1987 年版的《情报语言学基础》也将此篇文章作为一章收入,完善、丰富了情报语言学的学科体系。

比较研究法是张教授使用较为频繁的又一方法。他的许多文章都是从比较的角度对具体的语言性能进行分析,并从各种情报检索语言结合的角度提出弥补和改进的方法,例如:《体系分类法中

"集中与分散"的矛盾》、《情报检索语言中语词标识的功能与局限》、《体系分类法中的模糊现象及其消除方法》,等等;此外,《情报检索语言》一书在论述到每种语言时也都设有专节进行比较研究,如"标题法与体系分类法的比较","单元词法与标题法、组配分类法的比较","叙词法与单元词法、标题法、组配分类法、体系分类法的比较","自然语言与情报检索语言的比较",等等,这些都为我们对各种检索语言性能的认识提供了更加深入而具体的依据。

理想语言设计法是张教授对结构功能分析法的又一延伸,实际上是为了设计包含某些理想功能的情报检索语言工具而寻求各种结构支持的一种研究方法。"学科—事物概念组配型检索语言"便是他运用该方法进行新型语言设计的一次成功尝试,是他经过十多年的探索创制出的一种完全新颖的情报检索语言结构模式,具有深入的研究价值。

另外,其它研究方法在张教授科研工作的各个方面也都发挥着重要的作用。应当说明的是他并不是先认识了这些方法,再按照这些方法去研究,而是在具体的研究过程中创造了这些方法,因而这些方法并非生搬硬套,而是完全适合情报语言学研究需要的。这条从实践中寻找、总结方法,然后再返回指导实践的科研路线是很值得广大的科研工作者借鉴的。

4 拓展了情报语言学研究领域

80 年代之前,我国图书情报界对情报检索语言的研究面是有限的。据曾蕾统计[5],在 1949~1981 年的 32 年中,有 73.48%的文献在讨论分类法,而关于标题法、单元词法、叙词法、关键词法、引文索引法和索引工作的文献一共只占 12.71%;在研究分类法的文献中,体系分类法占 59.28%,而研究组配分类法的只占 8.9%,论及分面分类法的就更少了。然而仅仅依靠综合体系分类法是无法满足情报检索需求的,而且情报检索语言研究的单一化也不利于体系分类法的完善和发展,所以张教授认为应该扩大检索语言的研究范围。不仅需要研究分类法,还要研究主题法;不仅需要研究体系分类法,还要研究各种组配分类法;不仅需要研究人工规范语言,还要研究自然语言在情报检索中的应用问题。

张教授认为,分类法、主题法,作为情报检索语言的两大类型,都是为提高检索效率而创制的,它们的基本原理应该是一致的,所不同的是它们在表达主题概念及其间关系时采用的方式以及在满足情报检索需求时所使用的手段,因此必须将它们综合起来研究,这样不仅能找出蕴涵其中影响着情报检索效率的共同规律,还能通过对各种语言结构与功能的分析和比较,总结出每种检索语言的优点和不足,更便于相互学习和自我改进,甚至可以通过综合研究和相互借鉴,创制出具有更高检索效率的新型情报检索语言(分类法主题法一体化检索语言的创制便是这一思想应用成果的突出代表);他还指出组配分类法是体系分类法的发展,创制组配分类法的目的主要是为了克服体系分类法的列举式列类方法所造成的不能无限容纳概念的局限性,以及类目单线排列所造成的"集中和分散"的矛盾。于是,他将各种分类法和主题法汇总到"情报检索语言"概念之下进行统一研究,并在国内首次以"情报检索语言"为书名系统、全面地写著了一本综合论述各种情报检索语言的专著。

自然语言检索法主要研究自然语言在情报检索中应用问题,是情报语言学中的一个重要分支,也是我国情报检索领域的一个重要发展趋势。张教授早在 80 年代初就已经意识到了这一发展趋势的重要性,他在 1983 年发表的《论情报检索语言的研究、创制与普及》一文中写道:"关于扩大研究范围,似乎也应把对自然语言检索法的研究包括进去。"他认为自然语言检索系统和情报检索语言检索系统并不是完全独立的两个系统,事实上两者在很大程度上存在着优势互补的现象,应该将

两者摆到一起，统一研究。于是，他给自然语言检索和情报检索语言冠以“情报语言学”的总名称，并用“情报语言学基础”作为《情报检索语言》一书增订本的书名。《情报语言学基础》第二版更是将自然语言在情报检索中的应用作为一章纳入情报语言学研究的范畴体系，以反映近年来情报语言学学科的发展。这标志着情报语言学研究领域的进一步拓展和学科体系的进一步完善。

情报语言学是一门应用性的学科，张琪玉教授深知在这一领域中闭门造车是不现实的。正如他自己所说：“世外桃源的环境可能无助于一个学者的成功。”[6]他对情报语言学研究领域的拓展不仅表现在研究对象上，还表现在应用领域中。索引法便是情报语言学的一个主要应用领域。索引是开发、利用文献资源的一种重要工具，它在人们日常的研究、学习、工作乃至生活中都具有普遍的使用价值。然而索引学与情报检索理论、情报语言学、目录学、文摘学、图书编辑学等学科都有着千丝万缕的联系，它的涉及面很广，因而张教授认为索引学应当以索引这一事物的基本原理、基本方法和基本技术作为自己的研究重点，对各种类型的索引作统一研究，有些与其它学科有交叉而且看来与其它学科关系更加紧密的部分，则不必再作为研究重点。这一研究策略沿袭了他对情报语言学的研究思路，即力图在对各种研究对象的综合研究中探索其内在的共同规律，然后再将这一规律返回至每个具体的研究对象，分析其优点与不足，以期相互借鉴和改进，甚至于推陈出新。

档案检索是张教授对情报语言学应用领域的又一次拓展。为了使档案专业教育适应我国档案事业发展的需要，加强档案检索知识的教育，他主编了教材《档案检索》。这是情报检索语言系统知识和理论在档案检索中的一次成功的运用。书中全面、系统地讲授了档案检索系统的各个环节，分析了作为档案检索系统语言保证工具的情报检索语言，尤其是分类法和主题法的组织结构、应用形式和具体功能。

5 构建了情报语言学学科体系

张琪玉教授是情报语言学学科的开拓者，构建情报语言学学科体系是他在学术上的又一大贡献。到目前为止，情报语言学的学科建设大体可分为三个阶段，并分别以他专著的三个版本为标志。

1983年6月由武汉大学出版社出版的《情报检索语言》是张琪玉教授的代表性专著，被定为高等学校图书馆学和情报学的专业教材。该书独辟蹊径地在我国首次提出了以检索效率为中心的观点，并将等级体系分类语言、分析—综合分类语言、标题语言、单元词法、叙词语言以及关键词语言统一纳入情报检索语言体系，进行综合研究，以探索它们影响情报检索效率的共同规律和有效的改进途径，它的完成与出版为中国情报检索语言的研究打开了局面，同时也对研究工作的深入开展起到了推波助澜的作用，标志着我国情报检索语言框架的初步形成和情报语言学学科建设的起步。《情报检索语言》是我国情报检索语言教材中系统性、全面性和理论性都比较好的专著。范并思在《论图书馆学学科前沿的转移》一文中写道：“与图书馆现代化相比，检索语言研究有更多的创新意义。图书馆现代化研究是学习西方建立起来的领域，而检索语言研究却基本属于中国人自己创立的领域。在张琪玉以前，国外还没有人能在‘情报检索语言’的书名下将检索语言理论的内容讲述得如此系统而精密。”这一评价恰如其分地指出了这部专著的学术地位。

1987年11月，该书增订版易名为《情报语言学基础》，并作为中央电大图书馆学专业用书由武汉大学出版社出版。与1983年版相比该版增加了三章一节，分别是：“第一章 绪论——情报语言学一般问题”，“第九章 情报检索语言基本方法”，“第十一章 情报检索计算机化与情报语言学的发展”，并将第二章改名为“情报检索语言基本理论”，增加了“第四节 文献主题的构成因素及层次”。

这大大提高了情报检索语言的理论性，扩展和丰富了研究范围，并辅之以系统的研究方法，使得全书的全面性、系统性和理论性都进一步增强。本版的另一个重要特点便是引入了自然语言检索的问题，并与情报检索语言结合起来统一研究，在两者之上提出了情报语言学的概念。他指出"情报语言学是图书馆学和情报学的共同分支学科"，"是研究情报检索中语言保证问题的一门学科"，"其主要研究对象是情报检索语言，同时也研究自然语言在情报检索中的应用问题。"[7]这一举动标志着情报语言学作为一门学科已经建立，并步入世界学科之林。

1997年9月，《情报语言学基础》增订再版。本版在前版的基础上又增加了两章："第九章 分类法主题法一体化检索语言"和"第十章 自然语言在情报检索中的应用"，使情报语言学的研究更加全面和及时，并充分体现当今情报语言学的发展趋势，同时也标志着我国情报语言学学科体系的进一步完善和情报检索语言研究领域的进一步拓展。

6 新型情报检索语言的创制与未来情报语言学的探索

张琪玉教授在《世纪之交中国情报语言学发展之路》和《情报语言学领域亟待研究且潜藏较富的课题》两文中对未来情报语言学的研究课题作了明确的概括，这既表明了张琪玉教授自己对学术研究的开拓创新意识，同时也指明了我国21世纪情报语言学的发展方向。

他指出，情报检索语言的改进应主要落脚于提高其易用化程度，应加强对情报检索语言（尤其是体系分类法和叙词法）的分面化改造，注重情报检索语言与自然语言的结合，尤其是有关自然语言接口及对应转换词表的问题。

他曾在《情报语言学基础》第十一章中，概括说明了情报检索计算机化对情报检索语言的影响，指出情报检索计算机化促进了情报检索语言的创新和改造，使词表、类表机编化、机读化及编、管、用一体化，使检索方法有了很大提高，使情报检索语言兼容化和标准化，使情报检索语言的应用范围扩大，尤其是使自然语言在情报检索中的应用得以实现，并预见"情报检索计算机化今后将会更快、更广阔、更深入地得到发展"[8]。针对网上资源检索这一迫切而重要的研究课题，他认为要加强对适用于组织和检索网上资源的检索语言的探索，并指出应重视中外主要情报检索语言及自然语言之间的对应转换问题。

对于自然语言研究，张教授认为这无疑是个正确的方向，必须下大力气进行研究，以缩短与国外先进技术的差距和加快自然语言实用化的进程。但目前纯自然语言检索的水平还很低，而现实情报检索中并未出现放弃情报检索语言的迹象，因而他始终认为现在谈论纯自然语言检索时代还为期过早。他强调自然语言检索必须加以控制，必须将情报检索语言的原理运用到自然语言检索中去，比如他就提出自然语言检索中的大多数问题可通过采用后控制技术加以解决，而后控词表，按其原理和功能，也可被认为是一种情报检索语言。所以他认为，当前亟待从情报语言学角度深入研究自然语言，情报检索语言研究者应当积极参与自然语言检索法的研究。至于说到自然语言和情报检索语言的未来，他从情报语言学这一宏观的角度提出一种设想，认为"未来将是自然语言的情报检索语言化或情报检索语言的自然语言化"，"在这种新型情报检索语言普及以前的趋势可能是下列三种情况并存：情报检索语言与自然语言在一个检索系统中并用，情报检索语言增加自然语言成分，自然语言适当引进情报检索语言的原理与方法"[9]。

为探索21世纪的情报检索语言，张教授为自己心目中的理想情报检索语言划定了一系列标准：既可从学科、专业角度检索，又可从事物角度检索；既可按系统入手检索，又可按字顺入手检索；既可先组式使用，又可后组式使用；既可进行专指性检索，又可进行泛指性检索；既可用词进行标引

和检索,又可用号码进行标引和检索;既可用人工语言进行标引和检索,又可用自然语言进行标引和检索[10]。经过十年的艰辛探索,他认为目前构成这种理想情报检索语言的方法和技术已经存在,并终于运用理想语言设计法设计出了“学科——事物概念组配型检索语言”[11][12]。这是他对未来情报检索语言的一次摸索和尝试,是一种完全新颖的情报检索语言结构模式,具有深入的研究价值

张琪玉教授老骥伏枥,壮心不已,虽年届古稀,仍耕耘不止。衷心祝愿张琪玉教授取得更多更新的研究成果,也相信我国情报语言学将不断走向繁荣和发展。

参考文献:

[1]吴建中.21世纪图书馆展望——访谈录.上海:上海科学技术文献出版社,1996

[2]曾蕾.我国三十二年来检索语言文献的初步调查与简略分析.图书情报知识,1982(2)

[3]张琪玉.论情报检索语言的研究、创制与普及.图书情报知识,1983(4)

[4]同上

[5]同[2]

[6]俞君立.中国当代图书馆界名人成功之路.武汉:武汉大学出版社,1996

[7]张琪玉.情报语言学基础.武汉:武汉大学出版社,1987

[8]同[1]

[9]张琪玉.情报语言学基础.武汉:武汉大学出版社,1997

[10]张琪玉.探索21世纪的情报检索语言.北京大学学报(信息管理系专刊),1997

[11]张琪玉.学科—事物概念组配型检索语言——关于情报检索语言的遐想与求索.图书馆杂志,1997(2)

[12]同[10]

[13]张琪玉.情报检索语言.武汉:武汉大学出版社,1983

[14]张琪玉.档案检索.北京:书目文献出版社,1993

[15]张琪玉.情报检索语言方法综述.图书情报知识,1984(2)

[16]张琪玉.文献主题的构成因素及层次.图书情报知识,1985(1)

[17]张琪玉.情报检索语言中语词标识的功能与局限——关于主题法性能的几点分析.湖北高校图书馆,1985(1)

[18]张琪玉.关于《中图法》增加组配成分的可能性和方法的探索.北图通讯,1985(3)

[19]张琪玉.情报检索语言语法体系初探.图书馆理论与实践,1986(3)

[20]张琪玉.论后控制词表.图书情报杂志,1994(1)

[21]张琪玉.关于索引学研究和索引工作开展的设想和建议.江苏图书馆学报,1993(1)

[22]张琪玉.论索引项.图书馆杂志,1994(5)

[23]张琪玉.情报检索语言原理的一致和方法的差异.图书馆建设,1994(6)

[24]张琪玉.自然语言与人工语言对应转换——情报检索语言走向自动化之路.中国图书馆学报,1996(1)

[25]张琪玉.情报检索语言的发展趋势(与吴建中的对话).图书馆杂志,1996(4)

[26]张琪玉.情报检索语言中聚类的原理和方法.北京图书馆馆刊,1997(1)

[27]张琪玉.自然语言检索中各种因素对检索效率的影响.情报理论与实践,1997(5)

[28]张琪玉.世纪之交中国情报语言学发展之路.图书馆杂志,1997年理论学术年刊:151-157

[29]张琪玉.情报语言学领域亟待研究且潜藏较富的课题.图书馆杂志,1998年理论学术年刊:149-155

《情报语言学基础》评介

南昌大学信息管理系　彭在曦

[摘要]本文对《情报语言学基础》评介，认为张琪玉所著《情报检索语言》开创了系统地研究情报语言学的先河，修订易名为《情报语言学基础》进一步完善了情报语言学这门学科。主要是明确了情报语言学研究方向，丰富了情报语言学研究内容，独创了一套情报检索语言基本方法等。使我国情报语言学的研究进入到一个更高层次，并使我国情报语言学研究形成了一个新的格局。整个教材结构严谨，纵横比较，线索清楚，体现了著者优良的学风。但作为教材，如能配备一定数量的思考题和实习题，对于培养和提高学生分析问题和解决问题的能力，都会起到很重要的作用。

[关键词]情报检索语言　情报语言学　情报检索效率　图书评介

《情报语言学基础》一书是我国著名的图书馆学、情报语言学专家张琪玉教授优秀专著。熟悉情报检索语言的同志都不会忘记皮高品、刘国钧、杜定友等老一辈学者在研究情报检索语言方面有着不可磨灭的拓荒之功，而李兴辉、白国应、刘湘生、丘峰、翟凤岐、侯汉清等在研究情报检索语言方面的丰硕成果也是名满天下的。但系统地、完整地研究和撰写专门的情报语言学，各类型情报检索语言包括自然语言综合在一起研究并正式命名，张琪玉教授所著《情报语言学基础》具有首创之功。《情报语言学基础》(原名《情报检索语言》)问世十多年来，印刷9次近10万册，无论对情报语言学教材的建设，培养学术人才，还是对情报语言学这门学科的建设和发展，都作出了重大贡献。对水平如此高的著作要作出全面准确的评介，显然力不从心，这里仅谈谈自己学习后的几点体会。

1　《情报语言学基础》进一步完善了情报语言学学科

1980年6月，张琪玉教授完成《情报检索语言》一书，并于同年在武汉大学开设《情报检索语言》课程，该书被列入教育部主持制订的《高等学校文科教材编选规划》于1983年出版。这是我国情报语言学学科建设的开始，学科体系框架基本形成。情报检索语言是对分类检索语言、标题词检索语言、单元词检索语言、叙词检索语言、关键词检索语言等进行统一研究，探索它们影响情报检索效率的规律和提高检索途径的一门学科。

1987年电大版易名为《情报语言学基础》，增加了新的内容，即情报语言学一般问题、情报检索语言基本方法、情报检索计算机化与情报语言学的发展。增强了全书的全面性、系统性、理论性，完善了作为一门学科的基本条件。

1997年修订再版《情报语言学基础》，又增补了“分类法主题法一体化检索语言”和“自然语言在情报检索中的应用”内容，反映了该著作具有了与本门学科发展相适应的科学水平。如果说《情报检索语言》一书开创了系统地研究情报语言学的先河，那么《情报语言学基础》一书则进一步完善了情报语言学这门学科，使我国的情报语言学研究进入到一个更高的层次，并使我国的情报语言学研究形成了一个新的格局。

2 《情报语言学基础》为情报检索语言理论研究开辟了新路

2.1 明确了情报语言学研究的方向

具有关资料统计，在80年代以前，由于种种原因，我国的情报检索语言研究经历了相当一段时间的沉闷期，绝大部分的学术论文和著作都集中在体系分类法这一种情报检索语言类型，而且研究的重点花在体系分类法的思想性方面，忽视其学术性。拨乱反正后，尤其改革开放以来，重新起步的情报检索语言研究，进展速度之快是令人欣慰的，已逐渐把研究重点转移到如何改进和正确使用情报检索语言，以提高其检索效率的方面上来。但是，理论上还显得陈旧，缺乏深度，研究方向问题还未得到深人的探讨。在此情况下，1983年6月张琪玉教授所著的统编教材《情报检索语言》的出版，为情报检索语言研究明确了方向。书中把如何提高情报检索语言的检索效率问题贯串始终，非常明确指出，情报检索系统效率的高低，在很大程度上取决于所采用的情报检索语言的质量以及对它的使用是否正确。

2.2 丰富了情报语言学研究的内容

《情报检索语言》一书的出版，改变了我国情报检索语言研究中两个传统现象，一是长期以来绝大多数的论文几乎都集中在等级体系分类法方面，组配分类法和主题法系统中各种语言研究很少。二是把等级体系分类法和标题法(最早出现的一种主题检索语言)一直分开进行研究。书中明确指出，情报检索语言的一个发展趋势，是各种语言在方法上互相渗透，互相吸取，它们之间有共同规律，有必要对它们进行统一研究。这不仅扩大了研究范围，而且提高了研究水平。

《情报语言学基础》一书把自然语言在情报检索中的应用补充进去，阐述了电子计算机检索技术的发展，为自然语言检索开拓了光明的前景。并指出了自然语言检索法是否完全能取代情报检索语言，目前下结论为时过早。但情报检索语言与自然语言结合是情报检索语言发展的方向，这不仅丰富了情报语言学的内容，而且明确了情报语言学发展的正确方向。

2.3 创造了情报语言学研究方法

作为教材在传授知识的同时注意培养学生的能力，不仅使学生得到一筐“鱼”，而且掌握“渔”的本领。《情报语言学基础》第一章介绍了情报检索语言的研究方法，结构功能分析法、历史演进研究法、比较研究法等10种。在第十章特别补充了情报检索语言基本方法，详细阐述了从宏观的鸟瞰和微观的探索两个层次分析解剖情报检索语言结构与功能的关系，情报检索语言功能的完善，都是通过其结构的改进和增加来达到的。这是系统论结构功能法的巧妙运用，也是情报检索语言研究方法上的一种创造。

3 《情报语言学基础》结构严谨，纵横比较，线索清楚，适应教学

这部教材阐述了情报语言学的一般问题、情报检索的基本理论和方法，主要介绍了等级体系分类语言、分析综合分类语言、标题词型主题语言、单元词型主题语言、叙词型主题语言、关键词型主题语言、分类法主题法一体化检索语言和文献分析与标引，另外补充了自然语言在情报检索中的应用，以及情报检索计算机化与情报语言学的发展等。在结构安排上尽量体现了情报检索语言历史

发展过程，给人以学科的完整性。对于每一类情报检索语言，一般先论述其原理，然后介绍编制方法和使用方法，最后评述性能，指出优缺点，与其它检索语言在方法上互相吸取的情况。对于全书，先写绪论，基本理论，然后介绍各种情报检索语言，最后一章讲情报语言学的发展。对几个关键性问题，即情报检索语言基本方法、文献分析与标引问题、自然语言在情报检索中的应用、情报检索计算机化专设一章，作出较为全面系统的论述。这样有分有合，纵横比较，线索清楚的结构，加上作者流畅的文笔，便于教师授课，也便于学生掌握知识。特别是情报检索语言基本方法、自然语言在情报检索中的应用、情报检索计算机化和情报语言学发展等问题，都是我国情报检索语言研究的热点问题，不但可以帮助读者复习全书，理清情报检索语言发展线索，并且有利于培养独立研究、独立思考的能力。

4 《情报语言学基础》一书体现著者优良的学风

《情报语言学基础》著作之所以获得成功，主要是著者以极其严肃的科学态度，长期以来把情报语言学的科学研究和教学实践、教学研究紧密结合起来的结果。

张琪玉教授长期担任信息管理系主任，热爱图书馆事业，热心于教育事业，严于治学，这种优良的学风和他的著作一样，都是我国图书情报事业、教育方面的宝贵财富，而《情报语言学基础》教材正是张琪玉教授毕生治学的结晶之一。

《情报语言学基础》既是一部专著，又是一部教材，配备一定数量的思考题和实习题也是一个重要的组成部分，对于帮助学生巩固和深化所学理论，培养和提高他们分析问题和解决问题的能力，都会起到很重要的作用。

参考文献：

[1]中国图书馆学会．情报检索语言论文选．北京：书目文献出版社，1990
[2]张琪玉．情报语言学基础．武汉：武汉大学出版社，1997
[3]张琪玉．情报语言学基础．武汉：武汉大学出版社，1987
[4]张琪玉．情报检索语言．武汉：武汉大学出版社，1983